



Universidad de Granada

E.T.S. INGENIERÍA INFORMÁTICA

**Departamento de
Ciencias de la Computación
E Inteligencia Artificial**

TESIS DOCTORAL

**Aplicación de la Lógica Difusa a la Obtención de Conocimiento para la
detección y prevención de fraudes**

Luis Antonio Cerda Leiva

Granada, Julio de 2008

Editor: Editorial de la Universidad de Granada
Autor: Luis Antonio Cerda Leiva
D.L.: En trámite
ISBN: En trámite



**Aplicación de la Lógica Difusa a la Obtención de Conocimiento para la
detección y prevención de fraudes**

memoria que presenta

Luis Antonio Cerda Leiva

para optar al grado de

Doctor en Informática

Julio de 2008

DIRECTORES

María Amparo Vila Miranda

Daniel Sánchez Fernández

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
E INTELIGENCIA ARTIFICIAL

E.T.S. INGENIERÍA INFORMÁTICA

UNIVERSIDAD DE GRANADA

La memoria titulada “Aplicación de la Lógica Difusa a la Obtención de Conocimiento para la detección y prevención de fraudes” , que presenta D. Luis Antonio Cerda Leiva para optar al grado de Doctor, ha sido realizada en el Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada bajo la dirección de los Doctores María Amparo Vila Miranda y Daniel Sánchez Fernández.

Granada, Mayo de 2008

Los Directores

Fdo. María Amparo Vila Miranda

Fdo. Daniel Sánchez Fernández

El Doctorando

Fdo. Luis Antonio Cerda Leiva

Agradecimientos

Al concluir este trabajo, mis primeras palabras de agradecimiento son para todas las personas que de una u otra manera lo hicieron posible.

En primer lugar a Amparo y Daniel quienes no solo han compartido sus conocimientos y dirigido este proceso, sino también me han brindado su amistad y gran calidad humana.

También quiero extender este agradecimiento a todo el gran equipo que se desempeña en Granada (profesores, administrativos y becarios), quienes generan un maravilloso ambiente para el desarrollo y crecimiento profesional y personal, con lecciones de vida como las que a diario entregan Miguel, Juan Luis, Nicolás y Jesús. Por supuesto también a Luis y José María en Jaén.

En el ámbito laboral, agradezco a quienes tuvieron un rol destacado como Aníbal, quien contribuyó al inicio de esta aventura y a Maruzzella quien permitió concluirla, y en este largo recorrido a Rodrigo, Daniel y Graciela, Luis Alberto y Ricardo, también a mis colegas Eduardo y Viviana y a mis alumnos Rodrigo, Andrés, Claudio y Marco.

No puedo terminar estas palabras sin mencionar a mi esposa y a mis hijos a quienes dedico este trabajo, no sólo por comprender mi ausencia en periodos difíciles sino también por el gran apoyo que siempre han sabido brindarme.

Índice General

CAPÍTULO 1 INTRODUCCIÓN	11
1.1. OBJETIVOS.....	12
1.2. ESTRUCTURA DE LA MEMORIA.....	13
CAPÍTULO 2 FUNDAMENTOS DE GESTIÓN	14
2.1. LAS DECISIONES EN LA GESTIÓN ESTRATÉGICA	14
2.2. EL MODELO Y LOS INDICADORES DE GESTIÓN	16
2.3. GMROI.....	18
2.4. CICLO DE CONVERSIÓN DE CAJA	19
2.5. GESTIÓN DE INGRESOS Y PRECIOS.	22
2.6. SURTIDO EFICIENTE (CATEGORY MANAGEMENT)	27
2.6.1. Fidelidad.....	29
2.6.2. Exclusividad de Uso.....	30
2.6.3. Índice de Cambio de Marca.....	31
2.7. RELACIÓN ENTRE SURTIDO EFICIENTE Y GESTIÓN POR CATEGORÍAS	31
2.8. DEFINICIÓN DE LA CATEGORÍA	32
2.8.1. Rol de la Categoría.....	32
2.8.2. Valoración de la Categoría.....	33
2.8.3. Tabla de Evaluación y fijación de objetivos.....	33
2.8.4. Estrategia de la Categoría.....	34
2.8.5. Táctica de la Categoría.....	35
2.8.6. Implementación de la Categoría.....	35
2.8.7. Revisión de la Categoría e Implementación.....	35
2.9. NIVELES DE MADUREZ DE LA INFORMÁTICA DE GESTIÓN	35
2.9.1. La Curva de Nolan.....	35
2.9.2. Madurez de la Informática de Gestión.....	37
2.10. RESUMEN DE FUNDAMENTOS DE GESTIÓN	38
FUNDAMENTOS DE REGLAS DE ASOCIACIÓN	39
3.1. REGLAS DE ASOCIACIÓN	40
3.2. REGLAS DE ASOCIACIÓN DIFUSAS	42
3.3. PROBLEMÁTICA QUE REPRESENTA LA EXTRACCIÓN DE CONOCIMIENTO MEDIANTE REGLAS DE ASOCIACIÓN.	44
3.4. DEPURACIÓN DE LAS REGLAS DE ASOCIACIÓN.	46
3.5. VALORACIÓN OBJETIVA DE LAS REGLAS DE ASOCIACIÓN	47
3.6. VALORACIÓN SUBJETIVA DE REGLAS DE ASOCIACIÓN	49
3.7. RESUMEN DE FUNDAMENTOS DE REGLAS DE ASOCIACIÓN.....	55
ÁMBITO DE ESTUDIO: LAS GRANDES TIENDAS.....	56
4.1. LA INDUSTRIA DE LA VENTA AL DETALLE.....	56
4.2. TENDENCIAS GENERALES EN LAS GRANDES TIENDAS	57
4.3. EL ÁMBITO DE ESTUDIO DE LAS GRANDES TIENDAS	59
4.4. EL MODELO ACTUAL DE NEGOCIOS Y LA CADENA DE VALOR	61
4.5. LAS PRINCIPALES PROBLEMÁTICAS DE LA VENTA AL DETALLE.....	63
4.6. RESUMEN DEL ÁMBITO DE ESTUDIO.....	65
METODOLOGÍA PROPUESTA.....	66
5.1. DESCRIPCIÓN DE LA METODOLOGÍA FBI.....	69
5.1.1. Determinar el Perfil de la Empresa y su Modelo de Negocios.....	70
5.1.3. Limpiar, depurar y completar los datos	78
5.1.4. Análisis exploratorio preliminar.....	79
5.1.5. Interpretación de resultados.....	79
5.1.6. Análisis Comprobatorio (Parte 1: Extraer Reglas).....	80
5.1.7. Análisis Comprobatorio (Parte 2: Depurar Reglas).....	82

5.1.8. Actualización de la Lógica de Negocios.....	83
5.2. RESUMEN DE LA METODOLOGÍA PROPUESTA	83
PROCESO DE GESTIÓN DE RIESGO	84
CASO: DETECCIÓN Y PREVENCIÓN DE FRAUDE EN TARJETAS CON CRÉDITO DE GRANDES TIENDAS.....	84
6.1. DETERMINAR EL PERFIL Y EL MODELO DE NEGOCIOS DE LA EMPRESA	85
6.2. IDENTIFICAR LAS VARIABLES CLAVES Y LAS FUENTES DE DATOS.....	87
6.3. LIMPIAR, DEPURAR Y COMPLETAR DATOS.	90
6.4. ANÁLISIS EXPLORATORIO PRELIMINAR	92
6.5. INTERPRETACIÓN DE LOS RESULTADOS	96
6.6. ANÁLISIS COMPROBATORIO (PARTE 1: EXTRAER LAS REGLAS)	97
6.6.1. Definición de las Etiquetas Lingüísticas.....	97
6.6.2. Extracción de las Reglas de Asociación Difusa.....	100
6.7. ANÁLISIS COMPROBATORIO (PARTE 2: DEPURAR LAS REGLAS).....	100
6.7.1. Aplicación de las técnicas de Medidas de Interés a las reglas obtenidas	101
6.7.2. Medidas de Interés Subjetivas.....	102
6.8. INCORPORAR LOS RESULTADOS A LA LÓGICA DEL NEGOCIO	105
6.9. RESUMEN DEL ANÁLISIS DEL PROCESO DE GESTIÓN DE RIESGO.	106
PROCESO DE GESTIÓN DE NEGOCIOS.....	107
CASO: PROCESO DE GESTIÓN ESTRATÉGICA DEL NEGOCIO	107
7.1. DETERMINAR EL PERFIL DE LA EMPRESA Y EL MODELO DE NEGOCIOS.....	107
7.2. IDENTIFICAR LAS VARIABLES CLAVES Y LAS FUENTES DE DATOS.....	108
7.3. LIMPIAR, DEPURAR Y COMPLETAR LOS DATOS.	116
7.4. ANÁLISIS EXPLORATORIO PRELIMINAR.	117
7.5. INTERPRETAR LOS RESULTADOS.	127
7.6. ANÁLISIS COMPROBATORIO (PARTE 1: GENERAR REGLAS)	129
7.6.1. Etiquetas lingüísticas.....	130
7.6.2. Búsqueda de Reglas de Asociación.....	146
7.7. ANÁLISIS COMPROBATORIO (PARTE 2: DEPURAR LAS REGLAS)	148
7.7.1. Análisis de Item frecuentes.....	148
7.7.2. Aplicación de Técnicas de Medidas de Interés a las reglas obtenidas: ..	149
7.7.2.1. Medidas de Interés Objetivas.....	149
7.7.2.2. Medidas de Interés Subjetivas:.....	150
7.8. INCORPORAR LOS RESULTADOS A LA LÓGICA DEL NEGOCIO	154
7.9. RESUMEN DEL PROCESO DE ANÁLISIS DE LA GESTIÓN DE NEGOCIOS	154
CONCLUSIONES Y TRABAJO FUTURO	156
8.1. CONCLUSIONES DE LOS FUNDAMENTOS TEÓRICOS.....	156
8.2. CONCLUSIONES DE LOS CASOS BAJO ESTUDIO.....	157
8.3. CONCLUSIONES GENERALES.....	158
8.4. EL TRABAJO FUTURO.	159
8.5. SÍNTESIS LOS OBJETIVOS DE LA PROPUESTA DE TRABAJO FUTURO	159
8.6. BENEFICIOS ESPERADOS DEL TRABAJO FUTURO	160
ANEXO A. MINERÍA DE DATOS.....	161
A1. APROXIMACIÓN A LA MINERÍA DE DATOS	161
A2. LOS MODELOS DE LA MINERÍA DE DATOS	163
A3. EL PROCESO DE EXTRACCIÓN DE CONOCIMIENTO	164
A4. LAS TAREAS DE LA MINERÍA DE DATOS.....	165
A5. LAS TÉCNICAS O METODOLOGÍAS DE LA MINERÍA DE DATOS	167
A6. LOS MÉTODOS Y ALGORITMOS.....	168
A6.1. Algoritmo ID3 y C4.5.....	169
ANEXO B LÓGICA DIFUSA.....	173
B1. COMPUTACIÓN FLEXIBLE Y LÓGICA DIFUSA, UNA APROXIMACIÓN.	173
B2. El Origen de la Lógica Difusa y la Computación Flexible	174
B3. El problema básico.....	175

<i>B3. Los conceptos de la Lógica y de la Lógica Difusa</i>	177
B4. LA LÓGICA DIFUSA	177
B5. DIFERENCIAS ENTRE LÓGICA DIFUSA Y TRADICIONAL	180
A10. EL CONCEPTO DE SUBCONJUNTO DIFUSO	185
B6. DEFINICIÓN DE CONCEPTOS RELACIONADOS CON LA EXTRACCIÓN DE CONOCIMIENTO	189
ANEXO C VARIABLES Y ETIQUETAS LINGÜÍSTICAS.....	192
<i>C1. Variables Lingüísticas</i>	192
<i>C2. Etiquetas Lingüísticas</i>	193
D. LA INDUSTRIA DE LAS GRANDES TIENDAS	195
D1. EL PERFIL DE LAS GRANDES TIENDAS EN CHILE	195
<i>D1.1. Falabella:</i>	196
<i>D1.2. Ripley:</i>	197
<i>D1.3. Almacenes Paris:</i>	200
<i>D1.4. La Polar:</i>	201
D2. LAS GRANDES TIENDAS EN CIFRAS	203
D3. LA INDUSTRIA DE LAS GRANDES TIENDAS EN CHILE	203
<i>D3.1. Tendencias generales de la venta por departamento</i>	211
<i>D3.2. Otras tendencias generales de la distribución comercial</i>	211
<i>D3.3. Clasificación de los formatos comerciales tradicional y actual</i>	214
<i>D3.4. Formato de autoservicio</i>	215
<i>D3.5. Las estrategias comerciales</i>	221
<i>D3.6. Estrategia de penetración en el mercado</i>	222
<i>D3.7. Estrategia de internacionalización</i>	223
<i>D3.8. Estrategia de integración vertical</i>	230
D4. LA ATRACTIVIDAD DE LA INDUSTRIA DEL RETAIL Y LAS GRANDES TIENDAS	234
E. PROCEDIMIENTO PARA DETERMINAR EL PERFIL DE LA EMPRESA, SU POSICIONAMIENTO DE MERCADO Y EL MODELO DE NEGOCIOS ADOPTADO.	236
<i>E1. Proceso para Determinar el Perfil de la Empresa</i>	237
<i>E2. Proceso para Determinar el Modelo de Negocios</i>	241
<i>E3. Formularios para registrar el Análisis de cada nivel</i>	245
<i>E4. Resumen del Procedimiento</i>	246
F. RESULTADOS DE LA APLICACIÓN DE LA METODOLOGÍA FBI AL CASO DE GESTIÓN DE RIESGO Y CONTROL INTERNO	247
G. RESULTADOS DE LA APLICACIÓN DE LA METODOLOGÍA FBI AL CASO DE GESTIÓN DE NEGOCIOS.	252
BIBLIOGRAFÍA	258
PUBLICACIONES EN GENERAL	258
PUBLICACIONES PERIODÍSTICAS	268
SITIOS WEB DE INTERÉS	268

Índice de Figuras

FIG. 2.1 PRÁCTICAS DE NEGOCIOS, FUENTE: JEAN-MARC FRANCOIS.....	17
FIG. 2.2 CICLO DE CONVERSIÓN DE CAJA, FUENTE: REVENUE MANAGEMENT EN RETAIL [RI05A].....	19
FIG. 2.3 MODELO DE GESTIÓN DE CATEGORÍA. FUENTE: COMITÉ DE MARKETING-MERCHANDISING [CM99]	31
FIG. 2.4 CURVA DE NOLAN.....	36
FIG. 3.1. JERARQUÍA SOBRE MEDIDAS DE INTERÉS, FUENTE: ADAPTADO DE [RM07].....	46
FIG. 3.2 APROXIMACIÓN PARA EL DESCUBRIMIENTO DE CONOCIMIENTO, FUENTE: LUI ET AL [LB99]	50
FIG. 4.1. VENTAS TIENDAS POR DEPARTAMENTO EN CHILE AL 2005.	56
FIG. 4.2. DIAGRAMA DE CONTEXTO DEL ÁMBITO DE ESTUDIO, FTE: ADAPTADA DEL SITIO WEB TOMAX.COM.....	60
FIG. 4.3 PROCESO DE NEGOCIOS DE LAS GRANDES TIENDAS.....	61
FIG. 4.4 PROCESO DE ENTRADA DE PRODUCTOS A TIENDA.	62
FIG. 4.5 PROCESO DE EVENTOS Y PROMOCIONES.	63
FIG. 5.1. ÁMBITOS RELACIONADOS CON LA METODOLOGÍA FUZZY BUSINESS INTELLIGENCE (FBI)	67
FIG. 5.2 SÍNTESIS DE LA METODOLOGÍA FUZZY BUSINESS INTELLIGENCE (FBI).....	69
FIG. 5.3. PROCESOS QUE COMPONEN LA METODOLOGÍA FUZZY BUSINESS INTELLIGENCE (FBI).....	70
FIG. 5.4 SISTEMA: MACROENTORNO, INDUSTRIA, EMPRESA.....	71
FIG. 5.5 COMPONENTES DE SISTEMA: MACROENTORNO, INDUSTRIA, EMPRESA.....	72
FIG. 5.6 CADENA DE VALOR. FUENTE: MICHAEL PORTER[PM92].....	74
FIG. 5.7. MODELO COMERCIAL: TRES PERSPECTIVAS ESTRATÉGICAS. FUENTE: ARNOLDO HAX[HA01]. 75	
FIG. 6.1 ACTUAL PROCESO DE TRATAMIENTO DEL DESCONOCIMIENTO DE COMPRA. FUENTE: ADAPTADO DEL SOFTWARE MONITOR.....	86
FIG. 6.2 PROCESO DE TRATAMIENTO DE IMPUGNACIÓN O DESCONOCIMIENTO DE COMPRA.....	88
FIG. 6.3 DISTRIBUCIÓN DE IMPUGNACIONES AÑO 2002.	93
FIG. 6.4 DISTRIBUCIÓN GEOGRÁFICA DE LAS IMPUGNACIONES	93
FIG. 6.5 DISTRIBUCIÓN DE LAS IMPUGNACIONES CON RESPECTO A EXPENDEDORAS DE COMBUSTIBLE.....	94
FIG. 6.6 DISTRIBUCIÓN DE LAS IMPUGNACIONES CON RESPECTO A LOS PUNTOS DE VENTA PROPIOS.....	94
FIG. 6.7 DISTRIBUCIÓN CON RESPECTO A TIENDAS Y EXPENDEDORAS DE COMBUSTIBLE	95
FIG. 6.8 DISTRIBUCIÓN DE LAS IMPUGNACIONES CON RESPECTO A LA EDAD	95
FIG. 6.9 DISTRIBUCIÓN DE MONTOS	95
FIG. 6.10 ETIQUETAS LINGÜÍSTICAS DE EDAD DEL CLIENTE.....	98
FIG. 6.11 ETIQUETAS LINGÜÍSTICAS DE ANTIGÜEDAD DE LA CUENTA.....	98
FIG. 6.12 ETIQUETAS LINGÜÍSTICAS DE PLAZO DE COMPRA.....	99
FIG. 6.13 ETIQUETAS LINGÜÍSTICAS DE MONTO DE COMPRA	99
FIG. 7.1 ACTUAL MODELO DE TOMA DE DECISIONES.....	108
FIG. 7.2 MODIFICACIÓN DEL PERFIL.....	112
FIG. 7.3 GESTIÓN DE LA CAPACIDAD DE EXHIBICIÓN.....	113
FIG. 7.4 ADAPTACIÓN DEL LAYOUT DE TIENDA.....	114
FIG. 7.5 FUENTES DE DATOS.....	115
TABLA 7.2 DESCRIPCIÓN DE LAS VARIABLES SELECCIONADAS.....	116
FIG. 7.6 VENTA TOTAL EN TIENDAS GRANDES.....	118
FIG. 7.7 VENTA TOTAL EN TIENDAS MEDIANAS.....	118
FIG. 7.8 VENTA TOTAL EN TIENDAS PEQUEÑAS	119
FIG. 7.9 VENTA HISTÓRICA 2004-2006	120
FIG. 7.10 VENTA HISTÓRICA X LÍNEA.	120
FIG. 7.11 COSTO HISTÓRICO POR LÍNEA.....	121
FIG. 7.12 CONTRIBUCIÓN HISTÓRICA POR LÍNEA.	121
FIG. 7.13 VENTA MUJER PRIMAVERA.....	122
FIG. 7.14 VENTA POR ÁREAS MUJER.....	123
FIG. 7.15 COSTO VENTA POR ÁREAS MUJER.....	123
FIG. 7.16 CONTRIBUCIÓN POR ÁREAS MUJER.....	124
FIG. 7.17 VENTA SPORT CASUAL PRIMAVERA 2006.....	124
FIG. 7.18 VENTAS POR PRENDAS MUJER SPORT CASUAL	125

FIG. 7. 19 CONTRIBUCIÓN POR PRENDAS MUJER SPORT CASUAL.....	125
FIG. 7. 20 VENTA POLERA PRIMAVERA 2006.	126
FIG. 7. 21 VENTAS POLERA MUJER	126
FIG. 7. 22 CONTRIBUCIÓN POLERA SPORT PRIMAVERA 2005.....	127
FIG. 7. 23 VENTAS POR TIPO DE TIENDA.	128
FIG. 7. 24 ETIQUETAS LINGÜÍSTICAS DE LA CARGA DE ARTÍCULOS EN EXHIBICIÓN.	130
FIG. 7. 25 ETIQUETAS LINGÜÍSTICAS DEL CICLO DE VIDA DEL PRODUCTO	131
FIG. 7. 26 ETIQUETAS LINGÜÍSTICAS DEL CICLO DEL EVENTO	132
FIG. 7. 27 ETIQUETA LINGÜÍSTICA DE DISPONIBILIDAD.	133
FIG. 7.28 ETIQUETAS LINGÜÍSTICAS DE ESTACIONALIDAD DE LA PRENDA.....	135
FIG. 7. 29 ETIQUETAS LINGÜÍSTICAS DE EXISTENCIA DE ÍTEMS EN LA TIENDA (STOCK)	137
FIG. 7. 30 ETIQUETA LINGÜÍSTICA PARA COMPARAR VENTA DIARIA (ROTACIÓN).....	138
FIG. 7. 31 ETIQUETAS LINGÜÍSTICAS SEGÚN NIVEL DE VENTAS DE LA TIENDA	142
FIG. 7.32 ETIQUETAS LINGÜÍSTICAS DEL VELOCIDAD DEL OPERADOR LOGÍSTICO	144
FIG. 7. 33 ETIQUETAS LINGÜÍSTICAS VELOCIDAD DE REPOSICIÓN DEL PROVEEDOR	145
FIG. 7. 34 ETIQUETAS LINGÜÍSTICAS DE LA VENTA.....	146

Índice de Tablas

TABLA 2. 1 ROL DE LA CATEGORÍA. FUENTE: COMITE DE MARKETING-MERCHANDISING [CM99]	33
TABLA 4. 1 PERFIL DE LAS GRANDES TIENDAS.	56
TABLA 6. 1 DESCRIPCIÓN DE ATRIBUTOS DE LA TABLA CLIENTE.	91
TABLA 6. 2 DESCRIPCIÓN DE ATRIBUTOS DE LA TABLA TRANSACCIONES.	91
TABLA 6. 3 DESCRIPCIÓN DE LOS ATRIBUTOS DE LA TABLA DE TRANSACCIONES.....	92
TABLA 6.4 EJEMPLO DE ANÁLISIS DE JERARQUÍA SEGÚN SIMILITUD DE ANTECEDENTE.	103
TABLA 6.5 EJEMPLO DE ANÁLISIS DE JERARQUÍA SEGÚN SIMILITUD DE LOS CONSECUENTES.....	104
TABLA 6.6 EJEMPLO DE ANÁLISIS DE JERARQUÍA SEGÚN SIMILITUD DE AMBOS LADOS.....	104
TABLA 6.7 EJEMPLO DE ANÁLISIS DE JERARQUÍA SEGÚN CONFORMIDAD DE LAS REGLAS.....	105
TABLA 6.8 EJEMPLO DE ANÁLISIS DE JERARQUÍA SEGÚN CONFORMIDAD DE LAS REGLAS.....	105
TABLA 7. 1 TABLAS UTILIZADAS EN EL EXPERIMENTO.	114
TABLA 7. 3 TIENDAS POR TAMAÑO Y UBICACIÓN	117
TABLA 7. 4 VENTAS POR TIPO DE TIENDA.	127
TABLA 7. 5 DATOS UTILIZADOS EN LOS EXPERIMENTOS X DIA Y POR SEMANA.....	146
TABLA 7. 6 DISTRIBUCIÓN DE FILAS POR CANTIDAD DE VENTAS POR DIA Y POR SEMANA	147
TABLA 7. 7 EXPERIMENTOS CON VENTAS DIARIAS.....	147
TABLA 7. 8 EXPERIMENTOS CON VENTAS SEMANALES	148
TABLA 7. 9 COMPARACIÓN DE ÍTEMS FRECUENTES CON EXPECTATIVAS DEL USUARIO	148
TABLA 7. 10 EJEMPLOS DE REGLAS MUY FUERTES	150
TABLA 7. 11 JERARQUÍA SEGÚN SIMILITUD DE ANTECEDENTE UM -> TG.....	151
TABLA 7. 12 JERARQUÍA SEGÚN SIMILITUD DEL ANTECEDENTE \neg SA -> VM.....	151
TABLA 7. 13 JERARQUÍA SEGÚN SIMILITUD DE LOS CONSECUENTES UM -> TG	152
TABLA 7. 14 JERARQUÍA SEGÚN SIMILITUD DE LOS CONSECUENTES IC -> KC	152
TABLA 7. 15 JERARQUÍA SEGÚN SIMILITUD DE AMBOS LADOS DE LA REGLA UM -> TG (COSTADO IZQUIERDO).....	153
TABLA 7. 16 JERARQUÍA SEGÚN SIMILITUD DE AMBOS LADOS DE LA REGLA UM -> TG (COSTADO DERECHO)	153
TABLA 7. 17 JERARQUÍA SEGÚN CONFORMIDAD DE LAS REGLAS \neg SA -> VM	154

Capítulo 1 Introducción

Las decisiones estratégicas en las empresas se vuelven cada vez más complejas, debido a las múltiples variables que ha incorporado la competencia sin fronteras.

En esta competencia globalizada, los ciclos de negocios son cada vez más breves, esto exige agilidad y rapidez para que las decisiones sean oportunas, todo ello aumenta las condiciones de imprecisión y la incertidumbre

Las fuentes de información para las variables que intervienen en las decisiones estratégicas presentan diferentes dificultades, entre otras, las causadas por la complejidad y el gran volumen de transacciones almacenadas en las bases de datos y por la imperfección que estos datos presentan para los fines de gestión.

A lo anterior se agrega, las barreras tecnológicas existentes en la mayor parte de las empresas, las que habitualmente, cuentan con tecnología computacional orientada a resolver las necesidades de la operación diaria, pero son precarias en cuanto a tecnologías informáticas de apoyo a la gestión.

Esta carencia tecnológica, se da paradójicamente, donde el impacto puede ser crítico para la sobre vivencia del negocio, nos referimos a los ámbitos de las decisiones estratégicas, en donde las decisiones se toman basadas fundamentalmente en la intuición del analista o experto en la materia, cuyo ciclo de actualización de sus paradigmas se ha vuelto lento en comparación con la velocidad de cambio que ha impreso la globalización.

Este estudio analiza la eficacia de las técnicas procedentes de la computación flexible para apoyar la toma de decisiones estratégicas, generando una metodología orientada a mantener vigente el conocimiento del negocio basada en la extracción de patrones desde las bases de datos transaccionales que la misma empresa genera a diario pero no utiliza con este propósito por no contar con metodologías eficientes.

El ámbito de estudio corresponde a la industria de la venta al detalle, en particular, a las grandes tiendas, para ello se ha trabajado con las cuatro empresas que lideran la industria en el Cono Sur de América.

En el ámbito y en el contexto de este trabajo, surge la oportunidad de desarrollar la metodología propuesta, que permite superar la dificultades del actual escenario de toma de decisiones estratégicas,

utilizando los grandes avances tecnológicos en la informática, en particular en la extracción de conocimiento desde grandes bases de datos (Knowledge Discovery Database, KDD), término acuñado por Fayyad et al [FU96].

Entre los diversos modelos y técnicas existentes en el KDD, se ha elegido las Reglas de Asociación [AT93], las cuales, por su afinidad con la forma de razonar del ser humano, presentan características muy apropiada para enfrentar este desafío y, dadas las condiciones de imprecisión e incertidumbre con la que se toman las decisiones, en particular en la gestión estratégica, las reglas de asociación difusa son las que proveen mejor resultado [DM03].

1.1. Objetivos

Proponer una nueva forma de abordar la toma de decisiones estratégicas, mediante la representación de información imprecisa y técnica de minería de datos. Planteamos el uso de técnicas; herramientas y procedimientos provenientes del ámbito de la computación flexible para obtener conocimiento útil que permita mantener vigente el modelo y las reglas de negocios. De manera más específica,

Obtener el modelo, las características y las restricciones tecnológicas que en el ámbito de las decisiones estratégicas presenta en el actual escenario competitivo, en este contexto se plantean los siguientes sub-objetivos: (i) determinar tipo y nivel de uso de sistemas información de apoyo a los proceso de toma de decisión estratégica, (ii) determinar la existencia de oportunidades para mejorar la eficiencia de los procesos de decisiones estratégicas aplicando los métodos, técnicas, y procedimientos provenientes de la computación flexible, (iii) recopilar información del conjunto de procedimientos y técnicas de la computación flexible y evaluarlas como alternativas de solución para cubrir la brecha de eficiencia.

Obtener una metodología capaz de adaptarse a las distintas exigencias y a los distintos escenarios tecnológicos presentes en la industria, con los siguientes sub-objetivos: (i) establecer cuales son las técnicas, métodos y los procedimientos que permiten mejorar la eficiencia de los procesos de toma de decisiones estratégicas de negocios, (ii) Aplicar a dos casos concretos las técnicas, métodos y procedimientos mejor evaluados para mejorar la eficiencia de los procesos de toma de decisiones estratégicas de negocios.

1.2. Estructura de la memoria

A continuación se describe la organización y contenido de cada capítulo de esta memoria:

En el capítulo 2 se lleva a cabo una revisión de los fundamentos de la gestión estratégica y de los modelos de negocios propios del ámbito de la industria bajo estudio.

En el capítulo 3 se abordan los conceptos de Reglas de Asociación utilizadas como técnica de minería de datos para emplearla como componente fundamente en la solución del problema.

En el Capítulo 4 se presenta la industria bajo estudio y los principales modelos de negocios que las empresas del rubro utilizan, se describen los procesos y principales problemas que enfrentan.

En el Capítulo 5 se plantea la metodología propuesta para resolver la problemática descrita en el capítulo anterior.

En el Capítulo 6 se aplica la metodología propuesta al primer caso que corresponde al área de la Gestión de Riesgo.

En el Capítulo 7 se aplica la metodología propuesta al segundo caso relacionado con la Gestión de Negocios.

Finalmente se dan a conocer las conclusiones y el trabajo futuro en el capítulo 8. Adicionalmente en el Anexo A se incluyen los conceptos básicos sobre Minería de Datos; en el Anexo B una introducción a la lógica difusa; en el Anexo C una aproximación a la Etiquetas Lingüísticas; en el Anexo D una mirada a las principales empresas que integran la industria bajo estudio; en el Anexo E se detalla un procedimiento para determinar el perfil de la empresa y el modelo de negocios; y finalmente en los Anexos F y G los resultados detallados de la aplicación de la metodología propuesta.

Capítulo 2 Fundamentos de Gestión

Para introducir el tema, a continuación, se revisan los conceptos básicos relacionados con las decisiones estratégicas, con los modelos de referencia utilizados para tomar las decisiones y también se explican los indicadores de gestión que permiten al analista del ámbito de estudio focalizar sus decisiones. Finalmente este capítulo se presenta una categorización de los distintos grados de madures presenta la Informática de Gestión en general y en la industria bajo estudio en particular.

2.1. Las Decisiones en la Gestión Estratégica

Como ya fue indicado en el capítulo anterior, el interés de este trabajo es apoyar la toma de decisiones en la gestión estratégica, en consecuencia, para iniciar nuestra definición de conceptos, revisamos que significa Gestión Estratégica y luego los aspectos conceptuales de las decisiones.

Según la Real Academia Española las definiciones de los términos Gestión Estratégica son:

Gestión: "Hacer las diligencias conducentes al logro de un negocio";

Estratégico: "Algo es estratégico, cuando posee una importancia decisiva para su desarrollo".

En este contexto, las decisiones bajo estudio, están orientada a mantener o mejorar la posición competitiva de la empresa, por tanto, el éxito o el fracaso de estas decisiones afecta los resultados de la empresa. En consecuencia, la Gestión estratégica requiere que se tomen decisiones adecuadas en el momento oportuno.

Una decisión según Verdegay[TE94, pag53], tiene múltiples significados, se puede interpretar como una resolución, una determinación, una disposición o incluso una sentencia, esto quiere decir, que el concepto de decisión depende del contexto en el cual se aplica, y el proceso de toma de decisiones puede ser diferente en cada contexto.

En el problema clásico de decisión se definen elementos básicos: el espacio de alternativas, el ambiente en el que plantea el problema, el conjunto de resultados o consecuencias.

El ambiente o conjunto de estados de la naturaleza, es un conjunto de factores ajenos a quien toma la decisión y su definición depende de las características que presente:

Ambiente de Incertidumbre, donde no sabemos nada acerca de cual será el estado de la naturaleza que podrá presentarse.

Ambiente de Riesgo, sobre un conjunto de estados de la naturaleza hay definida una distribución de probabilidad.

Ambiente de Certidumbre, conocemos exactamente cual será el estado de la naturaleza que se presente.

Una decisión estratégica es una decisión tendiente a cambiar o mantener la posición competitiva de una empresa en su industria (conjunto de empresas del mismo rubro). La posición competitiva de una empresa se determina por la participación de mercado o nivel de ventas que la empresa detenta en su respectiva industria.

Por lo general las decisiones estratégicas se toman con información imperfecta y en ambientes de incertidumbre.

La información que nos provee el mundo real no siempre es suficiente, correcta o completa lo cual da origen a distintos tipos de información imperfecta, a saber:

Incetidumbre: No existe suficiente información para evaluar la certeza o falsedad de una determinada proposición. Ej. "mañana lloverá".

Imprecisión: No se especifica un valor concreto de un dato sino un conjunto al cual pertenece y depende de la "granularidad" del dominio. Ej. "Juan tiene entre 20 y 30 años".

Vaguedad: el conjunto que se especifica no esta bien definido, existiendo vaguedad disyuntiva y conjuntiva.

Vaguedad disyuntiva: el dato real es el único valor del conjunto, Ejemplo: "Juan es joven".

Vaguedad conjuntiva: el dato real tiene como valor todo el conjunto. Ejemplo: "Juan habla idiomas comerciales".

Imprecisión e Incertidumbre: la afirmación es imprecisa y no estamos seguros que sean ciertos. Ejemplo: "hay cierta posibilidad que el paciente tenga un "leve" fallo cardiaco".

Tipos de información imperfecta.

Información Errónea, la información es diferente de la existente en el mundo real, el caso más común es la inconsistencia.

Información Incompleta, la información no describe completamente la realidad.

Información Imprecisa, se dispone de un conjunto de valores entre los que se sabe está el verdadero valor del dato, esta información puede ser información nula o información disyuntiva.

Información Incierta, no se tiene seguridad acerca de la certeza de un dato, ya sea por mala calidad de la información o poca fiabilidad de la fuente.

En búsqueda de soluciones que resuelva estas imperfecciones, diferentes autores han trabajado distintas soluciones como los modelos matemáticos de la Teoría de la probabilidad, la Teoría de la Evidencia, la Teoría de los Factores de Certeza y la Teoría de los Conjuntos Difusos, Zadeh [ZL65], siendo esta última la que se utiliza en este trabajo.

2.2. El Modelo y los Indicadores de Gestión

La industria bajo estudio es una de las que ha tenido profundas transformaciones a lo largo de las últimas décadas. En ella se han incorporado innovaciones tecnológicas como la venta en línea, la distribución y/o reposición de materia prima y/o productos de parte de los proveedores externos e "internos" (bajo nuevos modelos de gestión como "justo a tiempo"), desarrollando nuevos sistemas logísticos y de transporte, progresivos avances de la simbolización, codificación y lectura óptica, innovando en los puntos de venta y en las estrategias comerciales de mantener cautivo al cliente (fidelización), utilizando el financiamiento directo por medio de una tarjeta de crédito emitida por el mismo establecimiento (tarjeta corporativa).

Estas innovaciones en modelos de negocios, en tecnología, en el formato de tiendas, ocurridas en los últimos años han producido cambios muy importantes dentro del sector. Esta renovación de los estilos de gestión ha obligado a este sector industrial a revisar sus prácticas para mejorar los márgenes, las que resume muy bien la siguiente gráfica:

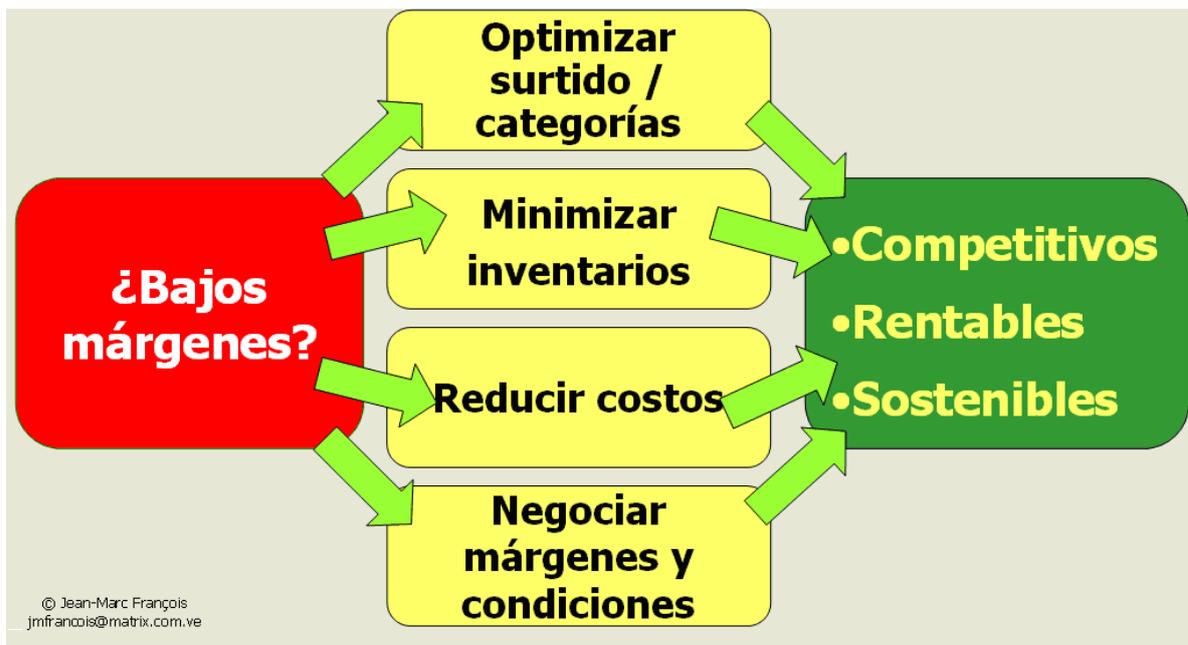


Fig. 2.1 Prácticas de Negocios, Fuente: Jean-Marc Francois

La esta figura 2.1, se resume las mejores prácticas que la industria ha consolidado y están asociadas a los modelos de gestión que requieren la incorporación de avances tecnológicos.

Uno de los modelos de gestión más recurrentes en esta industria es el modelo denominado Gestión de Ingresos y Precios (Pricing and Management Revenue, que se revisa en el apartado 2.6), el cual usa para su medición los siguientes indicadores de gestión:

Margen generado por inventarios: (Gross Margin Return of Investment, GMROI): es un indicador comparativo de desempeño de categorías, líneas de productos y gestión de inventarios.

Margen generado por metro cuadrado de superficie de ventas (Gross Margin Return of Space, GMROS): utilizado para comparaciones

cruzadas de ventas y categoría de productos (cross-store y cross-category)

Margen generado por empleado a tiempo completo (Gross Margin Return of Labor, GMROL): utilizado para comparaciones cruzadas de ventas y categoría de productos (cross-store y cross-category).

De estos indicadores el más utilizado es el GMROI que se revisa a continuación.

2.3. GMROI

De los indicadores de gestión descritos en el apartado anterior, el más utilizado en esta industria es el GMROI, en sus publicaciones autores como Sergio Flores [SF06] y Kim Russell et, al [KR99], entre otros, definen el Gross Margin Return on Inventory (GMROI) como el indicador que permite cuantificar la gestión de inventario de una manera más integral que el simple índice de rotación.

GMROI es el producto de dos índices que son la rotación y la contribución bruta sobre las ventas. Este indicador, como se observa de la expresión (1.1), conceptualmente es un índice adimensional que representa la rentabilidad obtenida por cada unidad monetaria invertida en inventario. Así, un producto o categoría de productos puede tener un valor alto de GMROI porque presenta un alto índice de rotación y porque genera una alta contribución por cada unidad monetaria vendida. Sin embargo, también es posible que existan productos con un alto GMROI como resultado de una baja rotación y una muy alta contribución.

$$\text{GMROI} = \frac{\text{Margen_Bruto}}{\text{Inventario_Promedio}} \quad (1.1)$$

El margen bruto se define como los ingresos menos los costos de la explotación; el inventario promedio son las existencias valorizadas que se tienen en promedio durante el período de tiempo en el que se mide el margen bruto.

Para incrementar el GMROI existen dos alternativas: incrementar el margen bruto con respecto al inventario promedio o bien disminuir el inventario relativamente más que el margen bruto. La primera alternativa produce otras tres alternativas, o se mejoran los ingresos

o se bajan los costos o se llevan a cabo ambas acciones simultáneamente. Dada la dificultad hoy en día de conseguir costos más bajos debido a la fuerte competencia e información de proveedores existentes, mejorar sustantivamente el margen bruto disminuyendo los costos de compra, que representan la mayor proporción de los costos operacionales, resulta difícil. Además en el corto plazo y asumiendo que el inventario ya está comprado, no hay muchas alternativas por esta vía más que planificar mejores compras en el futuro o hacer uso de un mejor poder negociador dado un mayor tamaño relativo de la empresa en el mercado. Por otro lado, aumentar los ingresos puede ser resultado de mejoras políticas de precio o de un mayor volumen de venta. El inventario promedio se reduce comprando más eficientemente y rotando más los inventarios.

2.4. Ciclo de Conversión de Caja

Tal como se plantea en el apartado anterior, el inventario promedio se reduce comprando más eficientemente. Además se consigue reducir el inventario promedio al mantener un menor nivel de inventario en el tiempo, lo que se logra comprando con mayor frecuencia aquello que se vende más rápido. La disminución del inventario promedio tiene efectos inmediatos principalmente en índices de actividad, rotación, endeudamiento y cobertura.

Es conveniente hacer referencia al ciclo de conversión de caja, concepto que ilustra en términos de tiempo, las etapas necesarias por las que pasa un producto desde que es comprado al proveedor, luego almacenado, exhibido y hasta que finalmente es vendido y se recupera la inversión al recibir el pago por parte del cliente final.

Se puede ilustrar el ciclo de conversión de caja de una empresa con la siguiente figura:

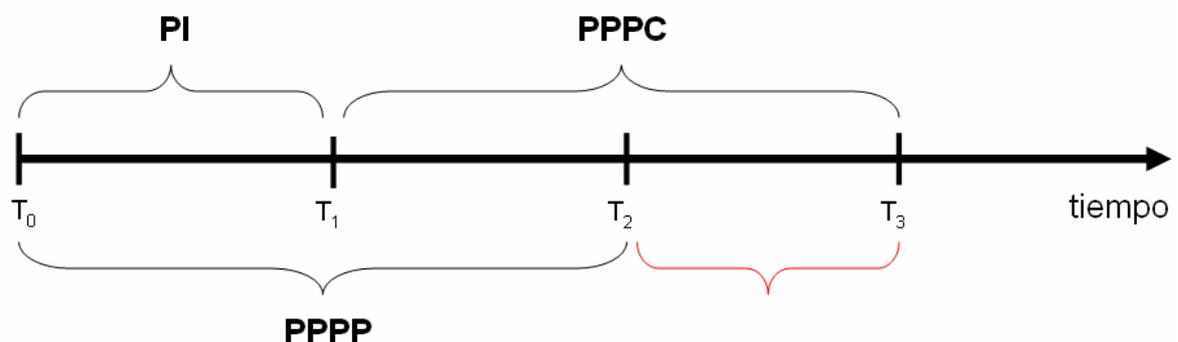


Fig. 2.2 Ciclo de Conversión de Caja, Fuente: Revenue Management en Retail [RI05a]

T_0 : momento en el que la empresa realiza la compra de un producto al proveedor

T_1 : momento en el que la empresa vende el producto al cliente

T₂: momento en el que la empresa debe pagar al proveedor por el producto

T₃: momento en el que el cliente paga a la empresa por el producto

El GMROI busca reducir la permanencia de inventarios (PI), con lo que se necesitaría menos tiempo de financiamiento para la compra del producto.

El ciclo de conversión de caja se divide en tres etapas fundamentales:

Permanencia de inventarios (PI): es la cantidad de días que el inventario permanece en la empresa desde que es adquirido de un proveedor hasta que finalmente se vende a un cliente.

Se calcula de la siguiente manera:

CxV= Costos de las ventas

$$\text{Rotación_Inventario} = \frac{\text{CxV}}{\text{Inventario_Promedio}} \quad (1.2)$$

La expresión (1.2) entrega la rotación del inventario en veces durante un período que la empresa convierte el dinero invertido en inventarios en ventas. Se calcula haciendo el cociente entre los costos de las ventas obtenidos durante un período (típicamente un año) y el inventario promedio durante aquél período. Si luego se hace el siguiente cociente:

$$\text{PI} = \frac{365}{\text{Rotación_Inventario}} \quad (1.3)$$

Se obtiene la permanencia del inventario en días.

Período promedio de pago de los clientes (PPPC): es el crédito que la empresa da a los clientes entendido por el tiempo promedio que la empresa da a los clientes para que estos paguen por las compras que realizan.

Se calcula de la siguiente manera:

IxV = Ingresos por Ventas

CxC= Cuentas por Cobrar

$$\text{Rotación_CxC} = \frac{\text{IxV}}{\text{CxC_Promedio}} \quad (1.4)$$

Expresión que entrega la rotación en veces en un período que la empresa rota sus cuentas por cobrar en promedio. Se calcula haciendo el cuociente entre los ingresos por venta obtenidos durante un período (típicamente un año) y las cuentas por cobrar promedio durante aquél período.

Si luego se hace el siguiente cuociente:

CxP = Cuentas por pagar

$$\text{PPPC} = \frac{365}{\text{Rotación_CxP}} \quad (1.5)$$

Se obtiene el período promedio de pago de los clientes en días.

Período promedio de pago a los proveedores (PPPP): es el crédito entendido por el número de días que el proveedor otorga a la empresa para pagar por la mercadería, una vez que esta físicamente pasa a formar parte del inventario.

Se calcula de la siguiente manera:

$$\text{Rotación_CxP} = \frac{\text{CxV}}{\text{CxP_Promedio}} \quad (1.6)$$

Expresión que entrega la rotación en veces en un período que la empresa rota sus cuentas por pagar en promedio. Se calcula haciendo el cuociente entre el costo de las ventas durante un período (típicamente un año) y las cuentas por pagar promedio en este período. Si luego se hace el siguiente cuociente:

$$\text{PPPP} = \frac{365}{\text{Rotación_CxP}} \quad (1.7)$$

Se obtiene el período promedio de pago a los proveedores en cantidad de días.

Para optimizar el ciclo de conversión de caja se recurre a dos modelos:

- Gestión de Ingresos y Precios (Pricing and Revenue Management)
- Surtido Eficiente (Category Management)

Que se revisan en los siguientes apartados.

2.5. Gestión de Ingresos y Precios.

La Gestión de Ingresos y Precios (Pricing and Revenue Management), en opinión de Ariel Baños en [BA05], es un instrumento de gestión de negocios surgida en los Estados Unidos a fines de los años noventa, ha logrado demostrar ser eficiente para lograr una mayor rentabilidad en la empresa. El buen uso de este instrumento puede resultar más eficiente que la gestión de inventarios y la reducción de costos.

De acuerdo a lo publicado en [MN06] Management & Negocios en la Nueva Economía, la importancia de la percepción del precio en los consumidores o clientes se produce porque el precio genera distintas reacciones, alguna de las cuales se revisan a continuación:

Percepción de ahorro, existe una percepción de precio-oportunidad y completamente opuesta conocida como precio lleno o completo. Es de uso común encontrar precios con decimales en muchos casos casi absurdos como por ejemplo \$0.99, en la mente del consumidor se percibe claramente que \$9.99 no es mucho mejor que \$10.00. A pesar que representa sólo \$0.01, la psicología del comprador asocia los .99 como muy favorable, por ellos esta percepción de ahorro es una estrategia de precios muy utilizada.

Percepción de valor, otra estrategia muy utilizada es la de agrupar productos bajo un precio atractivo, los denominados "combos", otorgando al comprador la percepción que lleva algo muy conveniente incluso alguno de ellos sin costo, es decir, "gratis". Es otra estrategia de uso común porque brinda a los consumidores la percepción que están obteniendo mucho más de lo que están pagando, es decir, perciben que ganan al comprar por que solo pagan una parte y se llevan otros productos/servicios adicionales de su interés o beneficio por el mismo precio.

Percepción de las rebajas, una rebaja en el precio de venta, tienen un enorme poder de atracción, mientras mayor sea el descuento la percepción de ahorro y mayor valor obtenido por menos precio es muy poderosa.

Lo anterior muestra la importancia del precio como instrumento de gestión, cuando se habla de La Gestión de Ingresos y Precios, no sólo consiste en una rebaja precios, dado que en determinadas ocasiones es necesario ajustarlos hacia arriba, y esto se puede dar aún en condiciones de estabilidad de precios en el mercado o condiciones competitivas muy desafiantes, puesto que las oportunidades de realizar ajustes de precios existen y se deben estudiar de manera mucho más analítica y selectiva.

En rigor, el instrumento Precio trata de buscar "pequeñas oportunidades", a modo de ejemplo, en temas como:

Análisis Caso-a-Caso de transacciones: El precio facturado generalmente es diferente al precio percibido. Conceptos tan diversos como descuentos financieros, promocionales, por garantías, transporte, o plazos de pago especiales son, entre otros, responsables de esta percepción distinta. Utilizando herramientas que permiten analizar la fijación de precios micro-analíticamente, transacción por transacción, operación por operación, pueden detectarse numerosas oportunidades que a menudo son pasadas por alto por los análisis convencionales.

Micro-segmentación de clientes: La evidente realidad que no todos los clientes son iguales, es una premisa para un cuidadoso análisis que permite ajustar selectivamente los precios de manera de capturar una mayor porción del valor creado en cada uno de los segmentos que integran el mercado objetivo de la empresa.

Adecuación de la Oferta y Demanda: La demanda, y la oferta de muchos productos fluctúa de manera muy significativa en el tiempo, lo cual exige precios dinámicos. Adoptar un enfoque de fijación de precios que se ajuste a estas fluctuaciones puede reportar muchos beneficios para las empresas, particularmente aquellas que venden servicios, productos perecederos, sujetos a modas o temporadas.

Al hablar de la Gestión de Ingresos y Precios se está hablando de repensar las políticas de precios como una herramienta estratégica para capturar valor e incrementar la rentabilidad de la empresa. Las empresas se concentran generalmente en reducir sus costos, mejorar la calidad, vender más, y ser más eficientes. Estos esfuerzos son sin duda fundamentales, pero dichas acciones no aseguran por sí mismas que la empresa sea rentable. La fijación de

los precios es el desafío que permite traducir estos múltiples esfuerzos de la empresa en resultados.

Con todas estas consideraciones la Gestión de Ingresos y Precios surge como una nueva herramienta para lograr la diferenciación de la competencia y la maximización de los ingresos de las compañías, sin embargo, su implementación no puede dejar de considerar factores como los que se mencionan a continuación:

- Localización geográfica
- Nivel socioeconómico de la zona
- Nivel de competitividad
- Hábitos de consumo de los habitantes de la zona de influencia
- Compra media
- Etc.

Esto significa en términos generales que se debe tener en cuenta tanto los datos de mercado como los datos del contexto en el que esta inmersa la compañía o la parte de ella que será sujeta de la fijación de precios, se debe tener en cuenta consideraciones como las siguientes:

- Nivel de comoditización del mercado.
- Nivel de valoración del Consumidor a los atributos de precio y plazo.
- Nivel de infidelidad en la compra.
- Potencial de mercado existente.
- Nivel de dispersión de precios.
- Criterios de formación del precios (concepto clásico: costos teóricos + margen teórico; sin tener en cuenta la competencia para ver si los precios fijados son los óptimos).
- Nivel de dispersión de la red de distribuidores / clientes
- Nivel de participación de Mercado (Tomadores / Seguidores de Precios)

Determinadas las condiciones anteriores, también conviene tener en cuenta factores como el posicionamiento comercial que desea la compañía frente a sus clientes, como por ejemplo ser visto como el líder en precios, o el líder en calidad, o el líder en conveniencia, teniendo en cuenta también el grado de esfuerzo económico y operacional que significa implementar la nueva fijación de precios, recordando que el desafío es garantizar la competitividad y hacer una correcta fijación de precios, el propósito último es asegurar los más altos ingresos y márgenes.

Para esto existen distintas estrategias de fijación de precios, que requieren de información oportuna de la competencia y del comportamiento histórico de la demanda propia y de la industria.

Dependiendo de la disponibilidad de esta información la empresa puede adoptar en una primera etapa adoptar una estrategia de precios reactiva y una vez que cuente con el conocimiento (información + experiencia) pasar a una segunda etapa de proactividad.

Las estrategias reactivas de precios, consiste en estar atento a los cambios manteniendo una capacidad de reacción inmediata a las acciones de precios como cambios de tarifas, comisiones y regulaciones de la competencia, teniendo presente que debe monitorear los resultados de las acciones propias y de la competencia, algunas acciones que permiten reducir los tiempos de reacción son por ejemplo implementar políticas de igualamiento automático, igualar tarifas, igualar regulaciones, mantener diferencias de precios en los segmentos que lo soporten.

Las estrategias proactivas buscan incrementar el ingreso neto mediante acciones innovadoras, anticipando a la competencia con actividades como por ejemplo: actividades promocionales, combinación de productos (combos), pruebas de micro-segmentación para maximizar la rentabilidad, decisiones de maximización de beneficios y minimización de pérdidas.

Según lo publicado por Vivanco [VJM05], El objetivo básico de la Gestión de Ingresos y Precios es lograr la rentabilización de los inventarios a través de ciclos dinámicos de procesos, ya que no basta simplemente con aumentar la rotación de los inventarios, si el margen de ventas es bajo, tampoco es bueno un alto margen acompañado de una baja rotación. Para su aplicación se usan técnicas que permitan predecir el comportamiento de los clientes a un nivel microeconómico y así optimizar el precio y la disponibilidad de productos con el fin de maximizar en principio la utilidad o beneficio que los productos reportan a la empresa, pero bien podría buscarse otro objetivo, como es aumentar la venta (en dinero).

La Gestión de Ingresos y Precios busca la maximización de la utilidad de los inventarios, por lo tanto maneja en conjunto la variable precio, así como el inventario. Sus orígenes están ligados a la industria del transporte aéreo, en este caso el inventario consiste en el número de asientos disponibles en cada ruta. En esta industria llevan más de 40 años de constante desarrollo e importantes resultados. Por ejemplo: American Airlines estimó beneficios cuantificables debido a esta técnica por US\$ 1,4 billones entre los años 1989 y 1991. Durante los años 90 estos conceptos y técnicas se expandieron al mercado retail, donde se han realizado muchos estudios, algunos de ellos con buenos resultados, pero evidentemente existe mucho por hacer en el tema, ya que aún no se trata de una realidad extendida en la industria, como si lo fue para las líneas aéreas.

La Gestión de Ingresos y Precios está orientada a la optimización de los montos invertidos en inventario y la utilidad económica que de ellos se obtiene.

Para la mayoría de los participantes en el mercado minorista las promociones o liquidaciones son parte importante de su rutina y tiene un impacto significativo en su resultado económico, por ello la determinación de las variables que influyen en una correcta política promocional es un problema desafiante e importante. Tales variables incluyen el descuento o precio de venta, el stock inicial de productos de temporada o la frecuencia correcta de los eventos de promoción entre otros.

Investigadores como Stephen Smith y Dale Achabal[SS00a], llevan años investigando modelos y políticas que permitan maximizar la ganancia de distintos actores del mercado retail en los Estados Unidos. En su trabajo: Precios de liquidación y políticas de inventarios para cadenas de retail [SS96], el estudio está orientado a encontrar el precio óptimo en periodos de liquidación. Para ello considera el efecto que pueda tener el precio (o porcentaje de descuento), el nivel de inventario y la estacionalidad propia en la demanda del producto. Algunas consideraciones del modelo son:

- La tasa de ventas depende explícitamente del precio, variaciones de temporada y el nivel de inventario.
- La tasa de ventas disminuye con bajos niveles de inventario, pero no se ve afectada por altos niveles de éste. Esta consideración surge de la experiencia de las personas encargadas en las distintas cadenas de retail incluidas en el trabajo. Especialmente importante en ropa cuando se tiene una incompleta selección de tallas y colores.
- Otras características importantes en el desempeño como: la competencia, la incertidumbre en la demanda y el tiempo de descuento, no están explícitamente incorporados al modelo.

Existiendo numerosas formas funcionales para la respuesta de la demanda al precio, utiliza una forma multiplicativa con sensibilidad exponencial la variable precio. Este tipo de funciones ha sido analizado y probado por varios investigadores, entre otros: MacIntyre et al [MS99], Kalyaman et al [KK96], Agrawal et al [AN03], pero ninguno de ellos incluye variaciones en la tasa de venta por estacionalidades o la influencia del nivel de inventario en la demanda, que son puntos claves y son temas importantes en muchas aplicaciones de retail.

Smith, Achabal y McIntyre en su trabajo: Maximizando el beneficio de promociones periódicas en tiendas de departamentos [SS00], señalan al precio o nivel de descuento y efectos de estacionalidades como

variables conocidas por su importancia en la respuesta a promociones, no así la sensibilidad a la frecuencia de promoción. La intuición sugiere que la promoción reduce la venta de artículos competidores y también la venta en las semanas siguientes.

En este caso la función de venta depende de las mismas variables que el modelo anterior, pero no considera que altos niveles de inventario no aumentan la tasa de venta e incorpora el efecto de la frecuencia promocional, así por ejemplo mientras un descuento se puede ver como una aceleración en la tasa de venta, la repetición de una promoción es una caída en la tasa.

2.6. Surtido Eficiente (Category Management)

A continuación se revisa e incluye lo publicado por Comité de Marketing-Merchandising de AECOC [MC99], donde se indica que el surtido eficiente es un proceso de colaboración entre distribuidores y proveedores para determinar la oferta óptima de producto, en una categoría, que alcance la satisfacción del consumidor objetivo al mismo tiempo que mejora los resultados del negocio.

El surtido eficiente asegura que los clientes o consumidores encuentran las opciones que buscan a la vez que refuerza el desarrollo de productos nuevos y mejorados para satisfacer sus necesidades evolutivas.

Por categoría, se entiende el conjunto de producto-servicio que el consumidor percibe como interrelacionado en la satisfacción de su necesidad, a partir de esta definición, el gerenciamiento de categorías en una nueva filosofía de gestión que está transformando la comercialización de productos de gran consumo y que se basa en gestionar los productos agrupándolos desde el punto de vista del consumidor de una solución a una determinada necesidad. Además de un nuevo concepto organizativo que contempla la gestión de las categorías como unidades estratégicas de negocio, el gestor de una categoría tiene responsabilidad total sobre las decisiones de surtido, niveles de inventario, distribución del espacio en el lineal, promociones y compras.

De acuerdo a lo indicado por en [MC99], los beneficios de un surtido eficiente son:

- **Valor para consumidores:** productos adecuados a la ocasión, en el lugar, momento y formato más convenientes. De este modo, las decisiones están orientadas al consumidor. Este cliente se ve beneficiado con lo siguiente:

- Existe claridad en su elección.
 - Facilita que encuentre las opciones que quiere.
 - Mejora su satisfacción al eliminar los quiebres de stock.
 - Adapta el surtido a sus necesidades particulares.
 - Obtiene productos nuevos y mejorados para satisfacer sus necesidades evolutivas.
- **Coherencia** entre los consumidores objetivo, las estrategias, objetivos y roles de la categoría en el sentido de que la definición de surtido es consistente con las decisiones más estratégicas de la compañía. A la vez en un surtido eficiente se deben identificar artículos que pueden parecer improductivos, pero que son importantes para el consumidor objetivo. Finalmente y para garantizar dicha coherencia entre las decisiones y los objetivos de la categoría, las decisiones respecto a surtido serán cuantificables.
 - **Cadena de suministro más eficiente:** que conduce a corto plazo a retornos financieros superiores y, a largo plazo, a una infraestructura logística y de fabricación mucho más eficiente. La metodología de surtido eficiente proporciona la habilidad para retirar artículos con un riesgo y un impacto financiero negativo. Del mismo modo asegura de manera justificada que los artículos introducidos son productivos e incrementan las ventas y el beneficio de la categoría, especialmente si no son totalmente nuevos en el mercado.

Por otra parte el proveedor obtiene los siguientes beneficios:

- Un conocimiento mucho más profundo de las dinámicas del consumidor en una categoría.
- Un conocimiento mejor de los puntos fuertes y débiles de los competidores de la categoría
- Otra perspectiva para el desarrollo de nuevos productos, formatos, tipos de envases, etc.
- Una apreciación más precisa de la dinámica financiera de la categoría en el nivel de la tienda (por ejemplo la importancia de la rotación, el inventario de tienda, etc.)
- Un instrumento para mejorar el desarrollo de nuevos productos.

El Surtido eficiente tiene por finalidad última satisfacer las necesidades del consumidor-cliente, para ello se debe contar con información de los hábitos de consumo y comportamiento de compra en relación a la categoría de productos específica. Se puede recurrir a fuentes de información acerca de sus hábitos de consumo con la información que se pueda obtener del punto de venta como de investigaciones de mercado, y las encuestas directas a nuestros clientes y sus sugerencias, de tal manera de conformar un panel de consumidores.

- **Panel de Consumidores:** El panel de consumidores permite establecer perfiles de consumidor (tipologías por áreas/canal y comportamiento según estudios socioeconómicos), tasas de penetración en hogares, ratios de frecuencia de compra así como conocer el volumen por acto de compra, el posicionamiento frente a las marcas y frente a las diferentes empresas de distribución.

Estudios sobre el consumidor y, especialmente, de comportamiento en el punto de venta que se realizan en colaboración con el distribuidor ayudarán a establecer, en primer lugar, el árbol de decisión del consumidor, es decir, los criterios de decisión sobre la compra del producto a partir de las características del mismo (marca, tamaño, variedad, etc.).

Además, permite conocer la tipología del comprador (perfil en cuanto a edad, tamaño hogar, renta, etc), lo que compra, la percepción de la categoría en la cadena, la cantidad comprada (compra media y precio pagado), la frecuencia de compra (promedio de actos de compra), el lugar de compra (fidelidad a las cadenas) y el proceso de compra (árbol de decisión, influencia de la actividad promocional, compra cruzada, recuerdo de precios).

También se requiere contar con información que relacione al consumidor con los productos que forman la categoría, en particular, medidas como:

- Fidelidad a las marcas, variedades, tamaños, tipos, etc.
- Exclusividad de la compra
- Índice de cambio de marca

2.6.1. Fidelidad

Mide el porcentaje de necesidades anuales que tiene el consumidor de una categoría y que son satisfechas por una marca determinada. Para ello se usa un índice que relaciona la cantidad comprada de esa marca y la cantidad comprada de producto por los compradores de esa marca. Su expresión matemática es la siguiente:

$$\frac{\text{Cantidad comprada de la marca A} * 100}{\text{Cantidad comprada del producto}}$$

Cantidad comprada del producto

Así si, por ejemplo, 1 comprador medio de aceite compra 20 botellas al año, y la media de la marca A para ese comprador es de 5 botellas por año, se podría estimar que la fidelidad a la marca A es del 25 por ciento (5 sobre cada 20 que compra). Esta relación tendrá siempre valores superiores a cero, siendo 100 su valor máximo, representando una fidelidad máxima del 100% (que significa que el comprador de la marca no compra otra marca en el mercado).

Esto también es aplicable para subcategorías, como por ejemplo, una especialidad, un tamaño, un sabor, etc. según los casos.

El valor de esta medida de fidelidad en la decisión de surtido puede ilustrarse con un ejemplo: Supongamos que un distribuidor tiene espacio para solo un artículo más en el espacio de exhibición asignado y hay tres artículos compitiendo por el espacio, uno con una fidelidad del 80 por ciento, otro con una del 40 y otro más con el 20 por ciento. Si el resto de los factores son aproximadamente iguales, la elección del artículo con mayor fidelidad del consumidor reduce la oportunidad de que los consumidores fieles a la marca acudan a otra tienda para satisfacer sus necesidades en esta categoría.

2.6.2. Exclusividad de Uso

La exclusividad de uso está definida como el porcentaje de una marca, variedad o subcategoría de la base de consumidores que han comprado sólo Marca A (o variedad A, etc.) durante un periodo de tiempo específico (los últimos tres, seis o doce meses). Cuando la exclusividad de uso de una marca es del 30 por ciento, significa que el 30 por ciento de los compradores la compran y que no compran ninguna otra marca dentro de la categoría durante ese periodo.

La medida de exclusividad puede ayudar a identificar artículos por los que los consumidores se ven tan fuertemente atraídos que seguramente le harán posponer su compra o irán a otra tienda a buscar el artículo.

La sutil diferencia entre fidelidad y exclusividad puede explicarse del modo siguiente: Un artículo con una fidelidad determinada podría tener una base de consumidores formada por usuarios no exclusivos. Por lo tanto, un surtido que excluya este artículo probablemente incluiría otros artículos alternativos. Por el contrario, si un artículo tiene un 40 por ciento de exclusividad en la base de usuarios, significa que este 40 por ciento de sus compradores no han comprado ningún artículo de la competencia recientemente y que retirar este

artículo del surtido, probablemente conduciría a un número importante de estos consumidores exclusivos a buscar el producto en otra tienda.

Medidas como "fidelidad" y "exclusividad" pueden proporcionar perspectivas de interés en cuanto a la magnitud del riesgo que conlleva la retirada de un artículo o, por el contrario, la oportunidad asociada de introducirlo.

2.6.3. Índice de Cambio de Marca

Mide el porcentaje de los compradores de una marca que también compran otras marcas (o formatos, o tipos) específicas en la categoría. Esta medida permite al distribuidor valorar si los consumidores de un artículo, que está considerado para su retirada (o introducción) tienen otras opciones en el surtido (artículos que han comprado en el pasado reciente).

2.7. Relación entre surtido eficiente y gestión por categorías

La siguiente figura muestra el modelo de Gestión de Categoría y muestra que la gestión del surtido es una de las cuatro "tácticas" en este contexto:

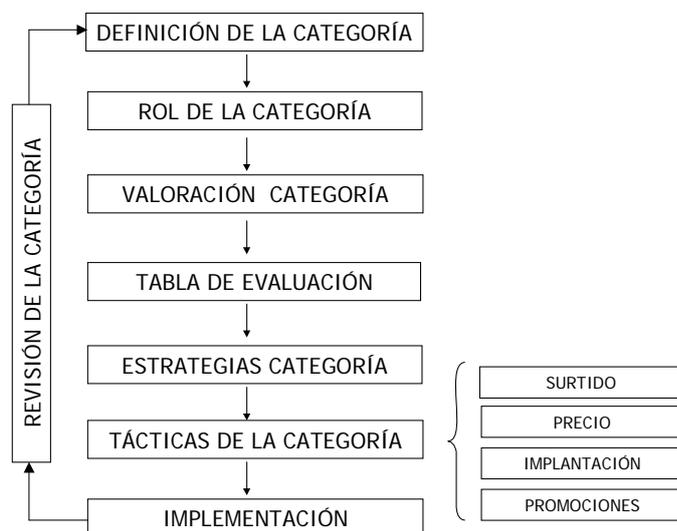


Fig. 2.3 Modelo de Gestión de Categoría. Fuente: Comité de Marketing-Merchandising [CM99]

A este respecto cabe recordar que el proceso de negocio de la Gestión por Categorías relaciona los objetivos y estrategias de la categoría con las acciones concretas para cada producto, formato o marca que lo componen.

De hecho, cada uno de los pasos del proceso de gestión por categorías tiene un efecto significativo sobre el surtido.

2.8. Definición de la categoría

La gestión por categorías define la categoría como "un grupo distinto y manejable de productos/servicios que los consumidores perciben como interrelacionados y/o sustituibles en la satisfacción de una necesidad".

En el proceso de definición de la categoría, en primer lugar, se identifican productos que ofrecen una solución parecida al consumidor (agua, cerveza, zumo, bebidas refrescantes, etc. "apagan la sed") y, en segundo, se seleccionan grupos más manejables que formarán la "categoría" (bebidas refrescantes, por ejemplo).

La segunda parte de la definición de categoría se refiere a la estructura u organización de la misma en subcategorías, también llamadas segmentos o necesidades. La estructura de la categoría describe como compran los consumidores aquella categoría y qué es importante para ellos según sus criterios de decisión.

La definición de la estructura de una categoría, que parte siempre de las necesidades del consumidor, nos ayuda a determinar las subcategorías correctas y los productos apropiados dentro de cada una de ellas.

El establecimiento de un surtido eficiente se basa en la estructura de la categoría que refleja una perfecta comprensión del comportamiento del consumidor. Primero se construye el surtido variedad por variedad (¿Queremos incluir o excluir esta necesidad específica del consumidor?). A continuación, se cuestionan las distintas opciones de necesidades del consumidor dentro de una variedad (¿Cuántos artículos ofrecer para cubrir esa necesidad?).

2.8.1. Rol de la Categoría

El rol escogido para una categoría, define su importancia sobre el conjunto de categorías del distribuidor o en los portafolios del

distribuidor y el proveedor. Cuando un distribuidor designa una categoría como "destino" le asigna más recursos, en este caso, un surtido más desarrollado, más amplio, que a una "de conveniencia".

Por lo tanto, un surtido eficiente para la misma categoría de un distribuidor A (para quien la categoría "X" es una categoría destino) debe ser diferente del surtido de un distribuidor B (para quien "X" juega un papel "de conveniencia").

Rol asignado a la Categoría	Cobertura relativa del surtido de esa Categoría
Destino	Normalmente la más alta
o Habitual	Moderadamente alta
o Ocasional- Estacional	Moderadamente baja excepto en la temporada "estacional"
o Conveniencia	Suele ser la más baja

Tabla 2.1 Rol de la Categoría. Fuente: Comité de Marketing-Merchandising [CM99]

2.8.2. Valoración de la Categoría

El proceso de valoración de la categoría analiza el rendimiento de la categoría a partir de cuatro perspectivas diferentes (el consumidor, el mercado, el distribuidor y el proveedor), cada una de las cuales tiene implicaciones para el surtido eficiente. Si, por ejemplo, el cliente objetivo del distribuidor concentra sus compras en una variedad determinada de la categoría (supongamos, tipo de fijación para cabello seca como la laca), un surtido eficiente debería proporcionar más opciones en esa variedad (más opciones en lacas que en gominas o espuma, siguiendo con el ejemplo).

Supongamos que el mercado muestra un crecimiento significativo en una variedad; bajo esas circunstancias, un proceso de surtido eficiente debería tener en cuenta este crecimiento y sugerir una ampliación del surtido.

2.8.3. Tabla de Evaluación y fijación de objetivos

La función de la tabla de evaluación es servir de norma para medir el éxito en la gestión de una categoría. Una tabla de evaluación debe incluir diferentes medidas internas (ventas, beneficio, etc.) así como externas (cuota de mercado, satisfacción del consumidor, etc.) para asegurar un equilibrio apropiado entre los distintos intereses operacionales y estratégicos.

Los indicadores escogidos en la tabla de evaluación pueden tener importantes implicaciones en la "eficiencia" del surtido:

- Se puede poner como ejemplo el caso en que uno de los objetivos para la categoría o variedad sea un aumento en el retorno en margen bruto sobre inversión. Satisfacer esta medida podría significar desviar el surtido de la categoría hacia más productos de entrega directa a tienda o marcas de elevado índice de rotación y alejarlo de productos que procedieran de plataforma (o almacén) o de marcas de baja rotación.
- De igual modo, el objetivo de aumentar el porcentaje de margen bruto de la categoría podría significar la eliminación de los artículos de menor margen bruto del surtido y añadir artículos de mayor margen, incluso si los artículos de menor margen son aceptables.

Un surtido difícilmente puede ser considerado "eficiente" si falla en el momento de satisfacer los objetivos de la categoría y las necesidades del consumidor. Por eso, la naturaleza de la tabla de evaluación de la categoría se convierte en un determinante importante dentro del contexto del surtido eficiente.

2.8.4. Estrategia de la Categoría

La misión de la empresa, sus estrategias y sus tácticas globales se reflejarán en las estrategias/tácticas a nivel de la categoría. Las distintas estrategias se definen según su impacto en variedades, marcas y artículos y pueden tener implicaciones en la eficiencia del surtido.

Por ejemplo, si una tienda A que es competidor clave de un distribuidor B se centra en un grupo específico de artículos como parte crítica de su surtido, este distribuidor puede "defenderse" añadiendo algunos de estos artículos al mix de sus productos. De igual modo, una estrategia de "generación de tráfico" debería incluir todos los artículos en segmento o necesidades con alta frecuencia de compra y elevada penetración en hogares. Una estrategia de imagen puede favorecer la inclusión en el surtido de formatos pequeños, de viaje, etc.

2.8.5. Táctica de la Categoría

La táctica de implementación de la categoría, consiste en el acuerdo entre fabricante y distribuidor sobre como se realizará su implantación y las condiciones en que se llevará a cabo.

2.8.6. Implementación de la Categoría.

La implementación simplemente consiste en la implantación de las decisiones tomadas, bajo las condiciones definidas en la etapa anterior.

2.8.7. Revisión de la Categoría e Implementación

Recogiendo las variables o indicadores utilizados para medir objetivos se comparan los resultados en cada variable una vez implantada la propuesta de surtido.

2.9. Niveles de Madurez de la informática de Gestión

La necesidad de caracterizar la relación de la Empresa con la Tecnología de Información es un esfuerzo que ha derivado en distintos modelos que emplean diferentes perspectivas para este propósito, de ellos hemos seleccionado la curva de Nolan para caracterizar la evolución teórica de la madurez informática y compararlo con la realidad verificada en las grandes tiendas a nivel de los sistemas de informática de gestión.

2.9.1. La Curva de Nolan

Propuesta por Richard Nolan en 1974, que estudió el comportamiento del gasto computacional versus tiempo, distinguiendo seis etapas.

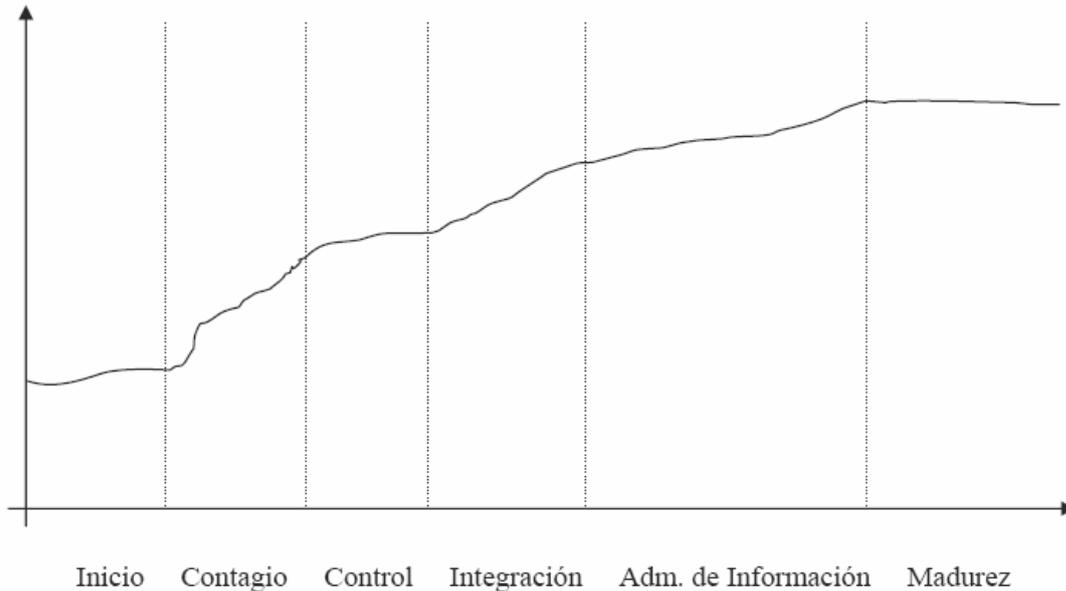


Fig. 2.4 Curva de Nolan

Contagio: los usuarios de información se dan cuenta de lo que esta herramienta puede hacer por ellos y solicitan el desarrollo de ciertas aplicaciones. El área de computación satisface estas solicitudes y comienzan a proliferar las aplicaciones computacionales. Todos desean recibir algún servicio computacional. El gasto crece en forma rápida y descontrolada.

Control: el desarrollo de nuevas aplicaciones y el marcado aumento del gasto en la etapa anterior llama la atención de la administración y pide detener el desarrollo de nuevas aplicaciones y reducir el gasto. Se reorganiza el área de Computación y se reestructuran algunas aplicaciones.

Integración: se definen políticas de desarrollo de aplicaciones. Se crea un sistema para contabilizar el gasto y en algunos casos se establece uno de cobro. Se introducen tecnologías que permitan integrar las aplicaciones desarrolladas en forma independiente. Crece el gasto pero en forma controlada.

Administración de la información: La organización se da cuenta que la información es un recurso y que las distintas áreas usuarias la comparten. Se desarrollan sistemas comunes a varias áreas. Dentro de la estructura organizacional el área de Computación se mueve a niveles donde pueda depender más directamente de Gerencia General o de Gerencia de Adm. y Finanzas.

Madurez: La mayoría de las aplicaciones de nivel Operacional y Táctico ya se desarrollaron, sólo queda generar sistemas para los niveles más altos. Se alcanza un grado de equipamiento más estable.

2.9.2. Madurez de la Informática de Gestión

Las expectativas que genera la incorporación de tecnologías de información distan mucho de los resultados observados en el nivel estratégico, esta realidad permite plantear la siguiente taxonomía de los niveles de madurez informática de gestión que coexisten simultáneamente en una empresa, en relación con lo planteado en el modelo de Nolan, la realidad de las grandes tiendas muestra que no todos los sectores ó áreas de la empresa alcanzan en forma simultanea el mismo proceso evolutivo como el descrito en el apartado 2.9.1. Esta clasificación considera que coexisten distintos grados de incorporación y uso de soporte tecnológico en los procesos de toma de decisiones, clasifica en seis niveles según sea el nivel de uso de tecnología y su efectividad en los procesos de obtención de información para la toma de decisiones, de esta forma sus categorías quedan como sigue:

Nivel Cero (No Informatizado): nula incorporación de tecnología de información, es decir, los procesos de obtención de información para la toma de decisiones se realiza por medios absolutamente manuales.

Nivel Uno (Semi Informatizado): uso incipiente de recursos tecnológico en los procesos de obtención de información para la toma de decisiones, por ejemplo el uso de planillas de cálculo como recurso de preparación de la información.

Nivel Dos (Informatizado Departamental): uso de sistemas de información administrativos (SIA) departamentales, que reúne los datos de los procesos asociados a una función específica dentro de la cadena de valor.

Nivel Tres (Informatizado Integrado): uso de sistemas de información administrativos integrados (ERP), los sistemas de información comparten los datos de todos los procesos que apoyan la cadena de valor de la empresa.

Nivel Cuatro (Informatizado Controlado o Sincronizado): uso de modelos de control de gestión como BSC integrados con los ERP, de manera tal que permiten tomar decisiones sobre la marcha al contar con información en línea del nivel de cumplimiento de las metas asociadas a los indicadores de gestión.

Nivel Cinco (Informatizado Predictivo): uso de modelos de minería de datos para extraer información no explícita que permiten determinar patrones de comportamiento y predecir tendencias para anticipar decisiones.

Nivel Seis (Informatizado Automático): uso de Sistemas Expertos con bases de conocimientos y motores de inferencia para apoyar la toma de decisiones.

Esta categorización del nivel de madurez de la informática de gestión en los procesos de toma de decisión permite clasificar el estilo de gestión de los procesos de decisión de las empresas en dos grupos:

Reactivas: se encuentran entre el nivel cero y el cuatro de madurez informática.

Pro-activas: se encuentran sobre el nivel cuatro de madurez informática.

La metodología que este trabajo presenta permite a las empresas cuyos procesos de decisión las clasifican en Reactivos al grupo de empresas que poseen procesos de toma de decisión estratégica Pro-activas.

2.10. Resumen de Fundamentos de Gestión

En este capítulo se han revisado los conceptos relacionados con las decisiones en la gestión estratégica basados en modelos de gestión asociados al ámbito bajo estudio y los niveles de madures de la informática de gestión.

De acuerdo a los antecedentes analizados, en el proceso de toma de decisiones estratégica bajo estudio destacará el uso del indicador de gestión GMROI. En cuanto al nivel de madures de la informática de gestión es predominante el nivel uno de madures de informática de gestión.

Capítulo 3

Fundamentos de Reglas de Asociación

La minería de datos, cuyo objetivo es, la obtención de conocimiento útil, no explícito desde los datos almacenados en grandes repositorios (Fayyad et al [FU96]), representa una metodología que ofrece una excelente alternativa de solución a la problemática bajo estudio, ya que como en otras áreas, es reconocido que la extracción de conocimiento es una necesidad básica, en nuestro caso específico, para determinar tendencias y patrones en una industria en constante evolución y con permanentes exigencias de decisiones rápidas y efectivas.

En esta tarea de obtener conocimiento útil no explícito desde grandes repositorios, se debe tener en cuenta, que es el usuario quien determina la utilidad del conocimiento, por ello se debe trabajar usando medios que se aproximen a su forma de razonar.

En el ámbito de estudio las decisiones se toman en un ambiente complejo y dinámico, donde la intuición es uno de los elementos fundamentales, en este sentido, la teoría de conjuntos difusos [ZL65], contribuye ciertamente a este propósito, dado que, es ampliamente reconocido que muchas de las relaciones que se dan en el mundo real son intrínsecamente difusas y las decisiones que se toman utilizan información incompleta o imprecisa.

En el contexto de la minería de datos, uno de los modelos mejor estudiados para la extracción de conocimiento es el de reglas de asociación [AT93], este modelo asume que el objeto básico de interés es un ítem, y que el dato aparece en la forma de conjunto de ítem llamado transacción. Las reglas de asociación son implicaciones que relacionan la presencia de ítem en las transacciones y es la transacción la estructura básica desde donde las reglas de asociación son obtenidas, en nuestros casos de estudio, las transacciones son las que se realizan en las grandes tiendas y también en el comercio asociado a estas grandes tiendas que acepta sus tarjetas de crédito como medio de pago.

A continuación se revisan los conceptos relacionados con la extracción de las reglas de asociación y con las medidas de interés que permiten depurarlas para discriminar dentro del gran volumen de patrones que esta metodología genera cuales son verdaderamente útiles e interesantes.

3.1. Reglas de Asociación

El propósito de esta técnica de minería de datos es buscar las posibles relaciones existentes entre la presencia o no de ítem en determinados conjunto de transacciones, por ejemplo se utilizan para identificar asociaciones entre la presencia o ausencia de artículos en transacciones comerciales, un ejemplo típico de regla de asociación es: el 90% de la gente que compra pan, compra leche, en lenguaje natural se puede expresar como: "la mayoría de la gente que compra pan compra leche", pudiéndose expresar mediante la notación $\text{pan} \Rightarrow \text{leche}$, en esta notación la parte izquierda se denomina antecedente y la parte derecha consecuente, a continuación se define formalmente el concepto de reglas de asociación.

Dado un conjunto de ítems, I , y un conjunto de transacciones T (al que también se le suele llamar T-set), donde cada transacción es un subconjunto de I , se dice que una regla de asociación [AT93] es una "implicación" de la forma $A \Rightarrow C$ que denota la presencia de los ítemsets (o conjuntos de ítem) A y C en algunas de las transacciones de T , asumiendo que $A, C \subset I, A \cap C = \emptyset$; y $A, C \neq \emptyset$.

Las medidas usuales propuestas en [AT93] para establecer la bondad y el interés de una regla de asociación son la *confianza* ($\text{Conf}(A \Rightarrow C)$), la probabilidad condicional $p(C|A)$ y el *soporte* ($\text{Supp}(A \Rightarrow C)$), la probabilidad conjunta $p(A \cup C)$.

Sin embargo, algunos autores apuntan algunas desventajas de la confianza, por lo que una solución alternativa es la que aparece en [BF01,02], donde se propone en su lugar el factor de certeza de Shortliffe y Buchanan [SE75]. El cálculo del factor de certeza de una regla de asociación se revisa en el apartado 3.5. Valoración Objetiva de las Reglas de Asociación.

Por definición, las reglas de asociación se encuentran definidas sobre conjuntos de transacciones. Dado que lo más común en una base de datos es trabajar con tuplas en lugar de transacciones, se han propuesto diversas soluciones a dicho problema. En caso de trabajar con bases de datos relacionales, lo usual es considerar que un ítem es cada par $\langle \text{atributo}, \text{valor} \rangle$ y que cada transacción se corresponde con una tupla en una tabla. Por ejemplo, diremos que el ítem $\langle A, a_0 \rangle$ está en la transacción asociada a una tupla t sí $t[A] = a_0$.

Sin embargo esta variante presenta algunos inconvenientes. En el caso de atributos numéricos, el número de ítems asociado a cada

posible par *(atributo, valor)* podría ser muy alto. Esta dificultad presenta dos problemas:

EL primer problema que presenta la extracción de conocimiento desde bases de datos relacionales, es la cantidad y calidad de las reglas de asociación extraídas, las medidas de soporte y confianza evalúan el interés y el grado de cumplimiento de las reglas de asociación con un enfoque meramente estadístico, si bien es cierto, el uso del soporte es bastante generalizado y se acepta como la mejor opción para medir la importancia, ya que reúne las siguientes ventajas [SJ03]:

- Es adecuado para la tarea de medir la relevancia estadística de una regla.
- Es una medida con significado intuitivo, y por tanto la interpretación de los valores de soporte es relativamente sencilla para el usuario.
- El uso del soporte contribuye al diseño de algoritmos eficientes para la búsqueda de reglas de asociación.

No ocurre lo mismo con la confianza, como medida del grado de asociación, implicación o dependencia entre antecedente y consecuente, ya que ha recibido diversas críticas. Los argumentos en contra de esta medida son los siguientes:

- La confianza no mide adecuadamente el grado de independencia estadística entre el antecedente y el consecuente.
- La confianza no refleja la dependencia negativa entre antecedente y consecuente.
- Por último, la confianza es una medida de probabilidad confeccionada. La probabilidad condicionada no es intuitiva, y por esta razón resulta difícil para un usuario no experto establecer umbrales mínimos de confianza semánticamente significativos a la hora de obtener reglas de asociación.

El segundo problema que presenta la extracción de conocimiento, surge de aplicar los modelos existentes a repositorios con presencia de atributos con dominios de un alto nivel de granularidad.

Este segundo problema que se presenta es el de la semántica de las reglas en sí y los parámetros que permiten al usuario interactuar con el proceso de extracción [DS99], sabiendo que el ser humano trabaja con descripciones del conocimiento con un menor nivel de granularidad, genera una dificultad de asignar un concepto claro al usuario, esto aumenta cuando se usa como criterio obtener reglas con el mayor soporte y confianza posible, aumentando además la complejidad del mecanismo de extracción.

En síntesis, por una parte, la alta fragmentación de las bases de datos, es decir, la presencia de atributos con alto nivel de granularidad, representa una incompatibilidad con la forma habitual del razonamiento humano y al intentar resolver este problema de incompatibilidad entre el alto nivel de granularidad y nuestra forma de razonar se generan otros problemas como el del mínimo soporte (*minsupp*), mínima confianza (*minconf*), tiempo de ejecución (*ExecTime*) y el aumento del número de reglas (*ManyRules*).

Algunas soluciones propuestas son las Reglas de asociación Generalizadas ([SR95]), donde se permite la definición de taxonomías entre los valores de atributos, o las Reglas de asociación Cuantitativas ([SR96],[ZL71]), donde a cada ítem se le hace corresponder un par $\langle \text{atributo}, \text{valor} \rangle$, esto es, se efectúa una partición en el dominio del atributo en intervalos. Otra posibilidad es la de dotar de un cierto grado de imprecisión a los valores, y extraer Reglas de asociación difusas, concepto que revisamos a continuación.

3.2. Reglas de asociación difusas

En la literatura, encontramos diversas propuestas de reglas de asociación difusas como una generalización de las reglas de asociación cuando los datos de partida son difusos o han sido tratados previamente para dotarlos de imprecisión [LJ97], [AW98], [KC98], [HT99], [DM03b]. Aunque la mayoría de estos enfoques se aplican sobre bases de datos relacionales, casi todas las medidas y algoritmos propuestos pueden emplearse en un marco más general. Se puede encontrar un interesante y completo estudio, con referencias bibliográficas sobre las extensiones a los casos de atributos cuantitativos o jerarquías de ítems, en [DM03a]. Otros enfoques adicionales al problema son los aparecidos en [ChK97], [KM02], [YH02], entre otros. En [DD03], se estudian diversas medidas para minería de datos difusos.

En [DM03], se incluye la definición de transacciones difusas como subconjuntos difusos de ítems, y será la idea en la que nos apoyaremos más adelante. Así, sea $I = \{i_1, \dots, i_m\}$ un conjunto de ítems y T' un conjunto de transacciones difusas, en el que cada transacción difusa es un subconjunto difuso de I . Dada la transacción $\tilde{t} \in T'$, notaremos por $\tilde{t}(i_k)$ al grado de pertenencia de i_k en \tilde{t} . Una regla de asociación difusa es una implicación de la forma $A \Rightarrow C$, tal que $A, C \subset I$ y $A \cap C = \emptyset$.

Resulta inmediato que el conjunto de transacciones donde aparece un cierto ítem es un conjunto difuso, al que se denomina *representación* del

ítem. Para el ítem i_k en T' , tenemos el siguiente subconjunto difuso de T' :

$$\tilde{\Gamma} i_k = \sum_{\tilde{\tau} \in T'} \tilde{\tau} (i_k) / \tilde{\tau}$$

Esta representación puede extenderse a un ítemset de la forma siguiente: Dado $I_0 \subset I$ un ítemset, su representación es el siguiente subconjunto de T_0 :

$$\tilde{\Gamma}_{I_0} = \bigcap_{i \in I_0} \tilde{\Gamma}_i = \min_{i \in I_0} \tilde{\Gamma}_i$$

Para medir el interés y la bondad de una regla de asociación difusa, se han de aplicar herramientas de razonamiento aproximado, debido a la imprecisión que puede afectar a las transacciones difusas. De acuerdo con lo expuesto en [DM03], se puede usar un enfoque semántico basado en la evaluación de sentencias cuantificadas [ZL83]. De acuerdo con [CJ95], Q es un cuantificador difuso coherente si verifica las siguientes propiedades:

$$Q(0) = 0 \text{ y } Q(1) = 1$$

Monotonía: Si $x < y$, $Q(x) \leq Q(y)$.

Dado Q , un cuantificador difuso coherente,

- El soporte de un ítemset I_0 en el FT-set T' es igual al resultado de evaluar la sentencia cuantificada Q de T' son $\tilde{\Gamma}_{I_0}$,

- El soporte de la regla de asociación $A \Rightarrow C$ en el FT-set T' , $Supp(A \Rightarrow C)$, equivale a la evaluación de la sentencia cuantificada Q de T' son

$$\Gamma_{AUC} = Q \cdot de \cdot T' \cdot son \cdot (\tilde{\Gamma}_A \cap \tilde{\Gamma}_C)$$

- La confianza de la regla de asociación $A \Rightarrow C$ en el FT-set T' , $Conf(A \Rightarrow C)$, se corresponde con el resultado de la evaluación de la sentencia cuantificada $Q \cdot de \cdot \tilde{\Gamma}_A \cdot son \cdot \tilde{\Gamma}_C$.

El factor de certeza, $CF(A \Rightarrow C)$, podemos seguir calculándolo de la misma forma en que lo hacíamos en el caso convencional no difuso, a partir del soporte y la confianza.

Las anteriores sentencias pueden evaluarse, por ejemplo, mediante el método GD, definido en [DM00a] como

$$GD_{\rho}(G/F) = \cdot \sum_{\alpha_i \in \Delta(G/F)} ((\alpha_i \cdot \dots \cdot \alpha_{i+1}) \rho) \left(\frac{|(G \cap F) \alpha_i|}{|F \alpha_i|} \right)$$

donde $\Delta(G/F) = \wedge(G \cap F) \cup \wedge(F)$, siendo $\wedge(F)$ el conjunto de niveles de F , y $\Delta(G/F) = \{\alpha_1, \dots, \alpha_p\}$ con $\alpha_i < \alpha_{i+1}$ para cada $i \in \{1, \dots, p\}$. Suponemos que el conjunto F está normalizado. De no ser así, F se normaliza y el factor de normalización se aplica a $G \cap F$.

Según se extrae de [DM03], el método propuesto cuenta con la ventaja de ser una generalización del enfoque clásico para la estimación de la bondad de una regla de asociación.

3.3. Problemática que representa la extracción de conocimiento mediante reglas de asociación.

A continuación se revisan los problemas que presenta esta metodología para extraer y seleccionar las reglas útiles y novedosas.

En primer lugar, el problema de extraer reglas de asociación se puede dividir en dos subproblemas:

1. Generar todas las combinaciones de ítems, entendiéndolas como ítemsets, con un soporte por encima de cierto umbral, previamente definido, el llamado soporte mínimo (*minsupp*). Esas combinaciones suelen encontrarse en la literatura con el nombre de ítemsets frecuentes.

2. Dado un ítemset frecuente $Y = i_1; i_2; \dots; i_k; k \geq 2$, generar todas las reglas que contengan todos los ítems de ese ítemset. Para ello, se toman todos los subconjuntos no vacíos X de Y y se generan las reglas $X \rightarrow Y - X$ que cumplan que su confianza es mayor que cierto umbral al llamado confianza mínima (*minconf*). El valor de la confianza viene dado por:

$$\frac{\text{Soporte}(Y)}{\text{Soporte}(X)}$$

Como Y es frecuente, cualquier subconjunto suyo también será frecuente [AT93]), así se tendrá almacenado su soporte. Las reglas que se deriven de Y satisfarán la restricción del soporte, pues Y así lo hace.

El principal inconveniente que presenta la extracción con reglas de asociación, estriba en que, en bases de datos lo suficientemente voluminosas, los costes tanto en tiempo como en espacio pueden resultar inviables. Por un lado, para lograr nuestro objetivo hemos de trabajar con todos los ítemsets posibles. Si tenemos m ítems, esto quiere decir que habremos de considerar 2^m posibles ítemsets.

Afortunadamente, los algoritmos existentes, aplican técnicas heurísticas para reducir, en la medida de lo posible, el número de ítemsets que se considerarán, de acuerdo a la estimación de si podrán o no ser frecuentes.

Contando con algoritmos capaces de disminuir exigencias de tiempo de proceso y espacio de memoria, y haciendo más eficiente los procedimientos, aún así nos podemos encontrar con otros problemas, ésta vez, los asociados a la aplicación que les pueda dar el usuario final.

El conocimiento obtenido es muy dependiente del contexto al que pertenece la información contenida en la base de datos original. Por este motivo, suele ser conveniente y necesaria la intervención de un experto humano que pueda dar una interpretación de las reglas obtenidas, indicando cuáles de ellas son potencialmente útiles y cuáles no, debido por ejemplo a su trivialidad [LB97].

Pero la labor del experto humano puede verse entorpecida si el conjunto de reglas obtenido es demasiado amplio. Es por eso que, de cara a optimizar la obtención y posterior interpretación de reglas de asociación dentro de una base de datos se pueden establecer ciertas restricciones, como por ejemplo:

Restricciones sintácticas. Estas restricciones limitan los ítems que pueden aparecer en una regla. Por ejemplo, podemos estar interesados sólo en las reglas que tengan un ítem específico en el consecuente o en el antecedente. También se pueden realizar combinaciones de las restricciones anteriores (como especificar que las reglas tengan ciertos ítems en el antecedente y otros en el consecuente).

Restricciones de soporte. Podemos estar interesados sólo en las reglas cuyos ítems aparezcan en un porcentaje de las tuplas de T por encima de un soporte mínimo. Esto quiere decir que para que la información que nos da la regla tenga cierto peso es necesario que aparezca con cierta frecuencia en la base de datos.

Restricciones de cumplimiento. La confianza nos da una medida de la fuerza de una regla. Nos informa sobre la dependencia entre la

aparición del consecuente si aparece el antecedente. En general, interesa que la confianza supere un mínimo.

3.4. Depuración de las reglas de asociación.

En problemas del mundo real no es fácil hallar implicaciones como las del ejemplo típico expuesto en el apartado 3.1., que cumplan con total certeza en un conjunto de transacciones y que ocurran un número de veces que las hagan interesante de considerar, con este propósito Agrawal et al[AT93], proponen las medidas de confianza y soporte, Bing Liu et al[LB97], proponen técnicas para medir la utilidad y la inesperabilidad de las reglas.

Este desafío, lo que plantea es: encontrar las denominadas Reglas Interesantes, [LB97], y para que una regla sea interesantes debe cumplir al menos dos requisitos: debe ser novedosa y útil, estas dos condiciones permiten satisfacer las expectativas del usuario, expectativas que no son estáticas en el tiempo, ya que corresponden a un momento determinado y a un contexto dado por el usuario, lo cual es cambiante en el tiempo. La siguiente figura muestra una taxonomía de las distintas técnicas o enfoques para enfrentar esta problemática:

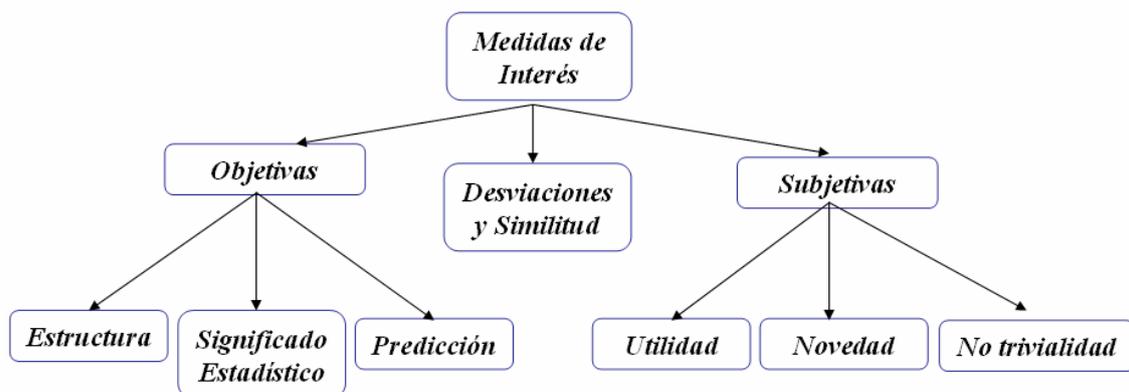


Fig. 3.1. Jerarquía sobre medidas de interés, Fuente: adaptado de [RM07]

Estos distintos enfoques para medir o cuantificar el nivel o grado de interés de una regla o patrón incluyen desde las medidas como el soporte, la confianza, fortaleza, rendimiento, significancia estadística, simplicidad, inesperabilidad y utilidad.

Las seis primeras son las llamadas medidas objetivas y se pueden aplicar sin la intervención usuario, las últimas dos son llamadas medidas subjetivas y requieren de su intervención.

3.5. Valoración Objetiva de las Reglas de Asociación

La figura 3.1, muestra tres de los aspectos más importantes tenidos en cuenta en el cálculo de medidas de interés objetivas: la estructura, el significado estadístico y la predicción.

En cuanto al uso de la estructura de la regla para medir objetivamente el interés se ha usado por ejemplo la frecuencia de apariciones de un ítem en las transacciones; en cuanto al significado estadístico se ha empleado el test de independencia y para la predicción se busca la capacidad de la regla para predecir otras.

Existen distintos estudios orientados a esta problemática, algunos de ellos son por ejemplo los trabajos Piatetsky-Shapiro et al [PG91], Berzal Et al [BF02], que buscan medidas de interés objetivas como por ejemplo en el caso de Berzal usan el Factor de Certeza de Shortliffe et al [SE75], en lugar de la confianza para determinar reglas fuertes y reglas muy fuertes, explicado en los siguientes apartado.

Factor de Certeza (CF): es una representación de la incertidumbre asociada al conocimiento, su principal objetivo es proporcionar unas medidas de incertidumbre más intuitivas que las medidas de probabilidad condicionada y sin algunos problemas de la teoría de la probabilidad, la certeza.

De un hecho en el intervalo discreto $[-1,1]$, donde la certeza de hecho (H) será:

1 si H es un hecho cierto;

-1 si H es un hecho falso y

0 si se desconoce el grado de verdad de H.

Berzal et al [BF01], se presenta la siguiente definición: el factor de certeza de una regla de asociación $I_0 \Rightarrow I_1$, se define como:

a) $Si \cdot Conf(I_0 \Rightarrow I_1) \gg Supp(I_1)$:

$$CF(I_0 \Rightarrow I_1) = \frac{(Conf(I_0 \Rightarrow I_1)) - Supp(I_1)}{1 - Supp(I_1)}$$

b) Si $\cdot Conf(I_0 \Rightarrow I_1) \setminus Supp(I_1)$:

$$CF(I_0 \Rightarrow I_1) = \frac{(Conf(I_0 \Rightarrow I_1)) - Supp(I_1)}{Supp(I_1)}$$

c) 0 en otro caso.

El factor de certeza toma valores en $[-1, 1]$, indicando el grado en el que nuestra creencia de que el consecuente es cierto varía cuando el antecedente es también cierto. Se mueve desde 1, el máximo incremento (esto es, si A es cierto, C también lo es), hasta -1, que indica el máximo decremento.

La nueva estructura que propone Berzal en [BF02], demuestra y define:

Para que una regla sea fuerte:

La regla $A \Rightarrow C$ es muy fuerte si $A \Rightarrow C$ y $\neg C \Rightarrow \neg A$ son fuertes:

- Condiciones de Soporte:
 1. $Supp(A \Rightarrow C) > Min\ sup$
 2. $Supp(\neg C \Rightarrow \neg A) > Min\ sup$
- Condiciones de CF:
 1. $CF(A \Rightarrow C) > Min\ sup$
 2. $CF(\neg C \Rightarrow \neg A) > Min\ sup$

Y agrega y demuestra lo siguiente:

Sea $(supp(A) + supp(C)) > 1$. Entonces $A \Rightarrow C$ es muy fuerte ssi $A \Rightarrow C$ es fuerte

Sea $(supp(A) + supp(C)) < 1$. Entonces $A \Rightarrow C$ es muy fuerte ssi $\neg C \Rightarrow \neg A$ es fuerte

Sea $(supp(A) + supp(C)) = 1$ entonces $A \Rightarrow C$ es muy fuerte ssi $\neg C \Rightarrow \neg A$ es fuerte

Tal como se muestra, el factor de certeza de una regla y su contabilizador de recíproco es el mismo cuando $Min\ cf \geq 0$.

Ejemplos de estos se aplican en los apartados 6.7 y 7.7 respectivamente.

3.6. Valoración Subjetiva de Reglas de asociación

Tal como se ha planteado anteriormente el proceso de minería extrae un gran conjunto de patrones, entre ellos patrones inesperados y novedosos, sin embargo, aunque un patrón sea inesperado o novedoso, eso no quiere decir que sea válido, incluso un patrón inesperado y válido puede ocurrir que no sea útil [LB93]. Para ello la utilidad del patrón o la aplicabilidad que la regla presente, depende del cumplimiento de expectativas del usuario, y tal como lo señala Sahar [SS00,01], no basta que la regla sea verdaderas sino que además debe ser interesante, entonces una regla puede estar en alguno de los cuatro siguientes grupos:

No verdaderas y no interesantes

Verdaderas y no Interesantes

No verdaderas e interesantes

Verdaderas e interesante

Naturalmente, las técnicas objetivas revisadas en 3.5, ayudan a determinar la veracidad de la regla, sin embargo, esto no es suficiente para determinar su utilidad, dado que además debe cumplir con otras condiciones para que un patrón o regla sea interesante, esto significa que requiere cumplir con las expectativa y con los fines u objetivos que el usuario requiere y esto depende exclusivamente de su ámbito y del momento que se encuentre viviendo, esta dependencia de las expectativas del usuario son dinámicas, es decir, pueden variar en el tiempo y están sujetas a las condiciones de contexto que el usuario enfrente en ese momento dado.

Para simplificar el proceso de encontrar aquellas que sean Verdaderas e Interesantes, diferentes autores [AR01], [BR99], [LB97, 99, 00], han acogido y desarrollado trabajos para determinar el nivel o grado de similitud entre los patrones obtenidos del proceso de extracción y aquellos que cumplen con las expectativa del usuario. Proponen por ejemplo establecer grados o niveles de jerarquía de las reglas según sus niveles de similitud. Al respecto, además de los criterios objetivos revisados en 3.2, se encuentran los aspectos subjetivos que derivan de la intuición o la creencia del usuario que

determinan el nivel la utilidad o de inesperabilidad de la regla extraída.

Para lograr este propósito Liu muestra en [LB99] la siguiente figura, en la que describe las distintas alternativas para enfrentar esta parte del proceso, estas van desde proveer directamente el resultado al usuario hasta proveerle una herramienta tecnológica que le asista en esta selección

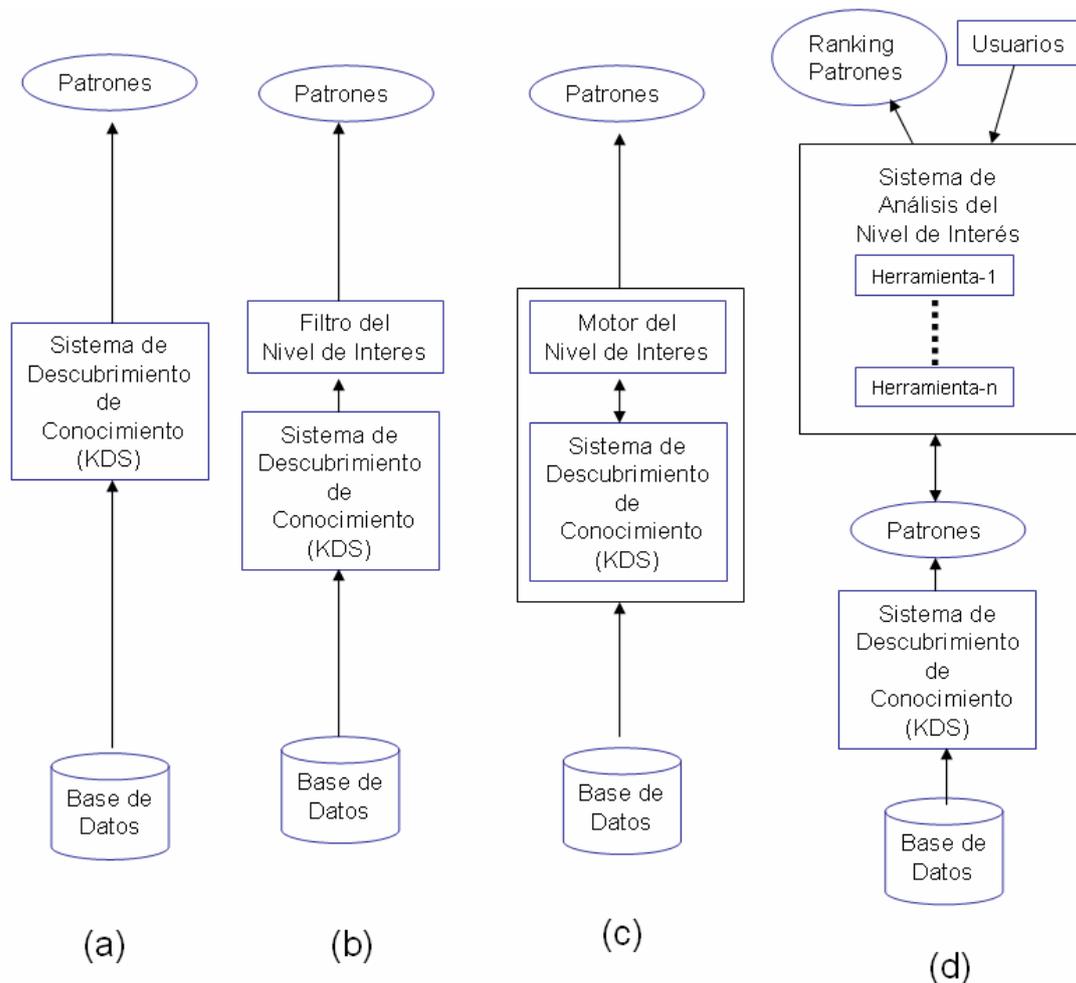


Fig. 3.2 Aproximación para el descubrimiento de conocimiento, Fuente: Lui et al [LB99]

En el modelo (a) se indica traspasar al usuario el resultado del proceso de extracción de reglas, con los inconvenientes de volumen de reglas que ello implica, en (b) se muestra cómo la búsqueda de patrones interesantes ocurre en un post-proceso a la minería. En (c) se muestra el proceso de búsqueda de patrones interesantes integrado al proceso de minería, y por último, en (d) se muestra una aproximación a la técnica formulada por Liu et al [LB99], proceso de post-análisis con una aplicación llamada Interestingness Analysis System (IAS) que ayuda al usuario a identificar patrones interesantes previo ingreso de las expectativas del usuario.

Como las medidas objetivas son insuficientes para determinar el grado de interés del patrón descubierto, algunos de los trabajos que usan desviaciones y medidas de similitud son por ejemplo los de Piatetsky-Shapiro y Matheus [PG94] quienes dirigen un agrupamiento de desviaciones de la normativa por medio de funciones de utilidad en un sistema KEFIR. También está el trabajo de Roddick y Rice [RJ01] quienes utilizan valores umbral basados en el comportamiento de los datos en el tiempo. Estos umbrales son modificados en el tiempo y según el contexto. El modelo toma en cuenta la importancia del evento y de la posibilidad de ocurrencia, por último mencionaremos el trabajo de Roussinov y Zhao [RD03], quienes desarrollaron el método CSSD (Context Sensitive Similarity Discovery) para detectar relaciones semánticas dentro de una aplicación de la minería de datos en el contexto de la Web, utilizando grafos con pesos en sus aristas, expresando las relaciones entre los vértices que representan a los distintos conceptos. Se construye mediante un proceso de agrupamiento jerárquico. Además CSSD utiliza técnicas de recuperación de información para crear los conceptos. Por otra parte, Bing Liu et al [LB99], propone una técnica denominada "User-Expectation Method", que utiliza en primer lugar la expectativa del usuario proveniente de su conocimiento anterior o histórico y/o sus sensaciones basadas en la intuición. Mediante una técnica de comparación difusa (fuzzy matching technique), revisa los patrones descubiertos con la minería de datos con aquellos proporcionados por el usuario, estableciendo una jerarquía de similitud.

Como se requieren medidas subjetivas, Liu propone una técnica para crear una jerarquía de los patrones descubiertos de acuerdo al grado de similitud, en particular el nivel de inesperabilidad.

En la técnica propuesta por Liu, lo primero es preguntar por las expectativas del usuario, luego un algoritmo de comparación que se utiliza para construir la jerarquía de los patrones descubiertos.

La técnica "User Expectation" propuesta por Liu, es una técnica interactiva de post-análisis, ayuda al usuario a analizar los patrones descubiertos.

En general es muy difícil si no imposible obtener completamente el conjunto del conocimiento existente en un momento particular del tiempo, esto se conoce bien como el sistema de adquisición de conocimientos (knowledge acquisition) en sistemas expertos.

Es probable que el usuario provea inicialmente un reducido conjunto de conocimientos y en procesos posteriores y recurrentes complemente este conocimiento. Sin embargo, este proceso incremental iterativo de proveer el conocimiento existente en el dominio del usuario no es siempre compatible o eficiente con algunas aproximaciones al problema dado que requiere que el usuario ingrese su conocimiento utilizando una técnica a fin con la técnica empleada para extraer los patrones de las bases de datos, como por ejemplo, las reglas de asociación.

La técnica propuesta para identificar los patrones útiles e inesperados esta compuesta por un proceso de dos pasos, en el primero el usuario provee un conjunto de patrones que espera encontrar, en el segundo paso un proceso compara los grados de similitud entre los patrones proporcionados por el usuario con los obtenidos en la minería de datos.

Al término de procesos como el propuesto por Liu, un patrón o regla será interesante cuando sea novedosa, útil y no trivial. Para que una regla sea novedosa depende del ámbito de referencia que se tenga, por ejemplo las creencias del usuario. Para que sea útil debe ayudar a conseguir algún objetivo del usuario y la no trivialidad depende de la autonomía del proceso y de su evaluación.

La técnica propuesta por Liu et al [LB97, 99, 00], para descartar reglas triviales o inútiles, señala una estructura para capturar y comparar los objetivos del usuario, por una parte están los componentes de las reglas entre los que destacan:

Los ítems o atributos situados al costado izquierdo de la regla (LI).

Los ítems o atributos situados al costado derecho de la regla (LD).

Los ítems o atributos situados a ambos lados de la reglas.

Los ítems claves o aquellos que deben aparecer siempre en la regla.

Por otra parte se encuentra el nivel de conocimiento previo que tiene el usuario acerca de las relaciones entre los ítems o atributos que componen la regla, el cual se puede dar según Lui de la siguiente forma:

El conocimiento general (cg), en este caso el usuario conoce cuales son los atributos relacionados, pero desconoce cómo se relacionan, es decir, si se encuentran al lado izquierdo (LI) o lado derecho (LD) de la regla.

El conocimiento razonablemente preciso (crp), el usuario conoce y sabe cuales de los atributos que pertenecen al costado izquierdo (LI) y cuales son del lado derecho (LD), pero no conoce con exactitud cuales atributos conforman la regla.

El conocimiento Preciso (cp), el usuario conoce la relación que existe entre los atributos que formarán la regla.

Al procesar el conocimiento previo del usuario y establecer sus niveles de similitud con el conjunto de reglas extraídas por el proceso de minería de datos se pueden obtener resultados como los siguientes:

Utilidad: la regla podría ser potencialmente útil si contiene sólo los ítems requeridos tanto para el lado izquierdo, lado derecho y ambos, lo cual no necesariamente la transforma en una regla interesante.

Inesperabilidad: una regla puede tener una connotación de inesperada, la que puede darse de diferente forma, por ejemplo:

Por tipo de atributo, la regla contiene atributos del tipo especificado por el usuario pero no son exactamente los indicados por el usuario, lo cual, requiere dedicar atención a estas reglas.

Por relación de atributos, la regla contiene los atributos especificados por el usuario, pero no cumplen con la relación que el usuario específico.

Atributos adicionales cumplen las medidas objetivas, atributos distintos a los especificados por el usuario se agregan a algunos de los atributos indicados por este y estas reglas cumplen con los umbrales indicados por el usuario.

Proceso de comparación de patrones

Sea:

E el conjunto de patrones de expectativas del usuario

B el conjunto de patrones descubiertos por el proceso de minería de datos.

W_i el grado de similitud entre un patrón descubierto $B_i \in B$ y el conjunto de patrones de E .

$W_{(i,j)}$ el grado de similitud entre un patrón descubierto $B_i \in B$ y el patrón de $E_j \in E$.

Cálculo de $W_{(i,j)}$ y W_i

1. Similitud por nombre de atributo

Se comparan los nombres de los atributos de las condiciones de B_i y E_j . El conjunto de nombres de atributos que son comunes a ambas condiciones de B_i y E_j es denotado por $A_{(i,j)}$.

2. Similitud por valor del atributo

Un atributo de B_i y E_j son comparados, las dos proposiciones son comparadas tomando en consideración los atributos de ambos operadores y sus valores.

Fórmulas para calcular $W_{(i,j)}$ y W_i

1. **Jerarquía de conformidad del patrón:** B_i y E_j tienen conformidad si ambas partes del condicional de B y E son similares.

2. **Jerarquía de patrones inesperados:** existen diferentes situaciones que permiten mostrar distintos tipos de patrones inesperados.

3. **Consecuentes inesperados:** la parte condicional de B_i y E_j son similares pero los consecuentes son distintos.

4. **Razones inesperadas:** los consecuentes son similares pero las partes condicionantes de B_i y E_j están claramente apartados:

Condiciones Inesperadas: el conjunto de atributos nombrados en el condicional de B_i y de E_j están solapados y son un poco distintos.

Atributos Inesperados: el conjunto de atributos usados en el condicional de B_i son inesperados con respecto de E_j .

3.7. Resumen de Fundamentos de Reglas de Asociación

En este capítulo se han revisado las reglas de asociación y las reglas de asociación difusas y las problemáticas que representa su extracción.

También se han revisado las técnicas propuestas para superar las problemáticas de extracción tanto con valoraciones objetivas como subjetivas, ejemplos de su aplicación se muestran en los casos de estudio que se incluyen en los capítulos 6 y 7.

Capítulo 4

Ámbito de estudio: las Grandes Tiendas

Este capítulo describe el ámbito de estudio y las principales problemáticas que presenta la toma de decisiones estratégicas.

Presenta una visión resumida de los participantes de la industria. La evolución histórica de cada uno de ellos se detalla en el anexo B.

4.1. La industria de la Venta al Detalle

La industria de la venta al detalle, también denominada Industria del Retail, en el Cono Sur de América es liderada por Chile, país que ha estado presentando durante los últimos años excelentes indicadores macro económicos. Esta sostenida favorable posición le ha permitido a Chile consolidar su economía y sus inversionistas locales han robustecido su liderazgo tanto nacional como internacional. En ese contexto, la venta al detalle es liderada por cuatro grandes tiendas como se puede apreciar en la siguiente gráfica:

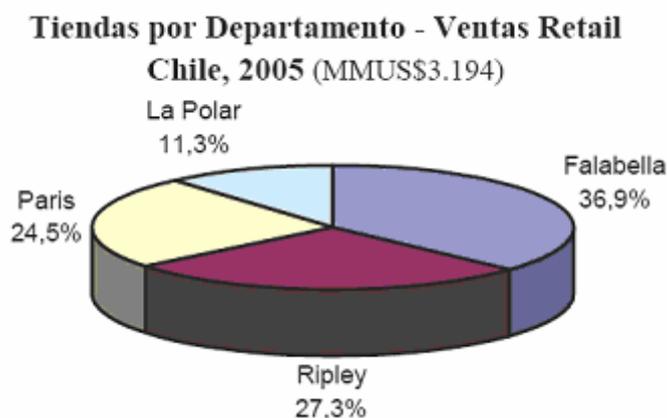


Fig. 4.1. Ventas Tiendas por Departamento en Chile al 2005.

El perfil de cada una de las grandes tiendas se resume en el siguiente cuadro:

Indicador	Falabella	Ripley	Paris	La Polar	Total
Nro. Locales	30	31	19	26	106
Superficie Venta m ²	177,538	227,909	154,544	81,000	640,991
Superficie / Locales	5,917	7,352	8,134	3,115	6,048
Ventas (MMUS\$)	1,178	871	782	360	3,194
% Mercado	36.90%	27.30%	24.50%	11.30%	100
Ventas/ m ²	3.55	2.6	2.6	2.3	
Tarjetas Emitidas (MM)	3.3	2.6	3.0	1.9	10.8
Tarjetas Activas (MM)	2.6	1.4	1.2	1.4	6.6
% Ventas con tarjetas	67%	63%	67%	80%	
Inversiones Proyectadas (MMUS\$)	1130	551	1200	100	

Tabla 4. 1 Perfil de las Grandes tiendas.

4.2. Tendencias Generales en las Grandes Tiendas

Los antecedentes para evaluar y analizar esta industria se incluyen en el anexo B, estos permiten apreciar que la tendencia de esta actividad económica es la aplicación de estrategias conducentes a mantener o mejorar su participación de mercado local y desarrollar nuevos mercados en el resto del Cono Sur de América, esto les permite concentrar un mayor poder de comprar. Las oportunidades de mejorar su negocio en el mercado nacional es cada vez más compleja debido a los altos niveles de competencia y a las similitudes de sus formatos comerciales. Como consecuencia, estas grandes empresas han explorado nuevos mercados en el extranjero donde la presencia de sus principales competidores nacionales no es tan fuerte y así han incrementado su cifra de beneficios. Claramente esta tendencia de internacionalización les permite conseguir un mayor poder de compra frente a mayoristas y fabricantes lo cual requiere de mucha precisión al momento de tomar las decisiones de compra, de manejo de inventario y de exhibición.

Además de las actuales tendencias en Chile de mejorar la masa crítica de clientes y que son tendencias generales del sector retail en el mundo globalizado, existen otras más específicas que deben tenerse en cuenta, como las que plantean Kotler en [KP00,99] y Vázquez en [VR97], las cuales se mencionan a continuación:

1. **El ciclo de vida de todos los formatos comerciales se acorta**, alcanzándose antes su madurez. Por ello, unos formatos sucederán a otros con mayor rapidez. Esto se debe a que cuando un formato comercial novedoso tiene éxito, rápidamente es imitado por otros, haciéndole perder enseguida ese carácter novedoso [KP00].
2. **La integración vertical de los canales de distribución** es un hecho creciente. Existe una tendencia hacia el acortamiento de los canales de distribución, en busca de alcanzar un control más efectivo de las funciones de distribución. Los canales de distribución excesivamente largos

son ineficientes debido a los mayores costes añadidos en los controles y a la dificultad de motivar a una larga red de intermediarios independientes [VR97].

3. Se está produciendo paulatinamente un ***aumento del poder de negociación de las grandes cadenas de distribución*** frente a los fabricantes. Gracias a que tienen sistemas de información superiores pueden utilizar información de marketing sofisticada y sistemas logísticos para suministrar un buen servicio e inmensos volúmenes de producto a precios muy atractivos para masas de consumidores. En este proceso, la gran distribución está excluyendo a los fabricantes más pequeños, que se ven dependientes de una gran empresa de distribución comercial y son, por tanto, extremadamente vulnerables. Muchas empresas de distribución comercial están incluso diciendo a los fabricantes más poderosos lo que tienen que hacer, cómo deben fijar sus precios, cómo deben hacer la promoción de sus productos, cuándo y dónde expedir la mercancía, e incluso cómo reorganizar y mejorar la producción y la gestión. En esta situación los fabricantes tienen pocas alternativas, ya que oponerse a las directrices de la distribución comercial podría suponer arriesgar entre un 10 y un 30 por ciento del mercado [KP00].
4. ***El merchandising es un factor diferenciador.*** El consumidor, que es cada vez más exigente, busca cada vez más comprar en un ambiente agradable, lúdico y sofisticado.
5. ***Los fabricantes se transforman en distribuidores.*** Son cada vez más numerosas los fabricantes que abren espacios de venta con su propio nombre con el fin de optimizar su marca. Un ejemplo de esta situación en Chile son las tiendas “*factory*” (Nike, Levi’s, etc.).
6. Cada vez más empresas adoptan el concepto de ***gestión integrada de la logística***, que se basa en una concepción

de la distribución física que reconoce la necesidad que tiene la empresa de integrar su sistema de logística con los de sus proveedores y clientes con el objetivo de maximizar los resultados de todo el sistema de distribución [KP99].

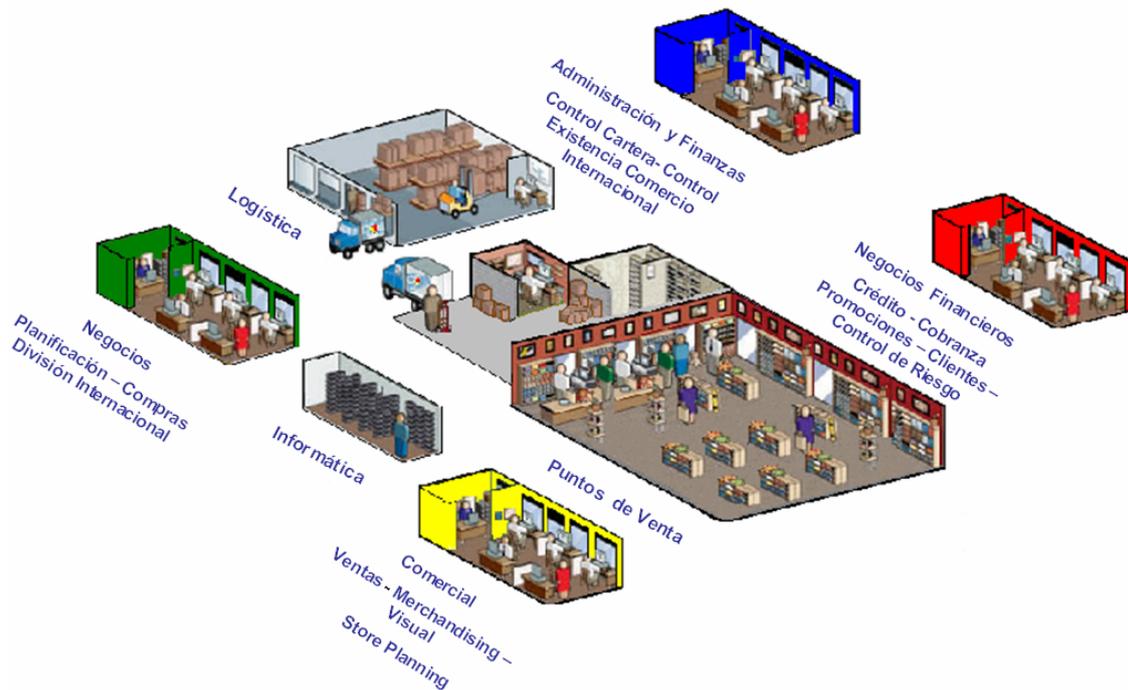
7. La ***demanda está cada vez más segmentada***, por lo que existirán muchas oportunidades para aquellas empresas que desarrollen una estrategia de nicho o de especialista.

8. Finalmente, la *era electrónica* ha acelerado de forma significativa el crecimiento de la venta sin establecimiento. Los consumidores reciben ofertas a través de sus televisores, ordenadores y teléfonos, a las que pueden responder de forma inmediata a través de llamadas gratuitas telefónicas o del propio computador [KP00]. Por tanto, la tendencia de la venta a través de nuevos canales es más que notable. Entre estos nuevos canales hay que destacar la gran importancia del **comercio electrónico**.

4.3. El ámbito de estudio de las Grandes Tiendas

El ámbito de estudio en el contexto de las grandes tiendas, en términos generales lo representan un diagrama que muestra la figura siguiente, adaptada del sitio Web TOMAX.COM en la cual destacan las áreas: Comercial, Negocios (Retail), Negocios Financieros, Logística, Informática, Administración y Finanzas y Tienda.

Diagrama de Contexto



» Fig. 4.2. Diagrama de Contexto del Ámbito de Estudio, Fte: Adaptada del sitio WEB TOMAX.COM

La actividad de las grandes tiendas se pueden agrupar en seis ámbitos: Comercial; Negocios Venta al Detalle; Negocio Financiero; Logística; Informática; y finalmente Administración y Finanzas. En el ámbito de Negocio, propiamente tal, se encuentran las áreas de Planificación, Compras y Comercio Exterior. En el ámbito Comercial se consideran las áreas de Marketing, Visual, Store Planning y Ventas. En el ámbito de los Negocios Financieros las áreas de Crédito, Cobranza, Promociones, Clientes y Control de Riesgo, del ámbito de Administración y Finanzas las áreas de Control de Cartera, Control de Existencias y Comercio Internacional, del ámbito de la Logística las áreas de almacenamiento y transporte y finalmente ámbito de la Informática.

4.4. El modelo actual de negocios y la cadena de valor

Para enfrentar las actuales tendencias competitivas las Grandes tiendas, recurren en general al proceso que gráfica en la figura siguiente:

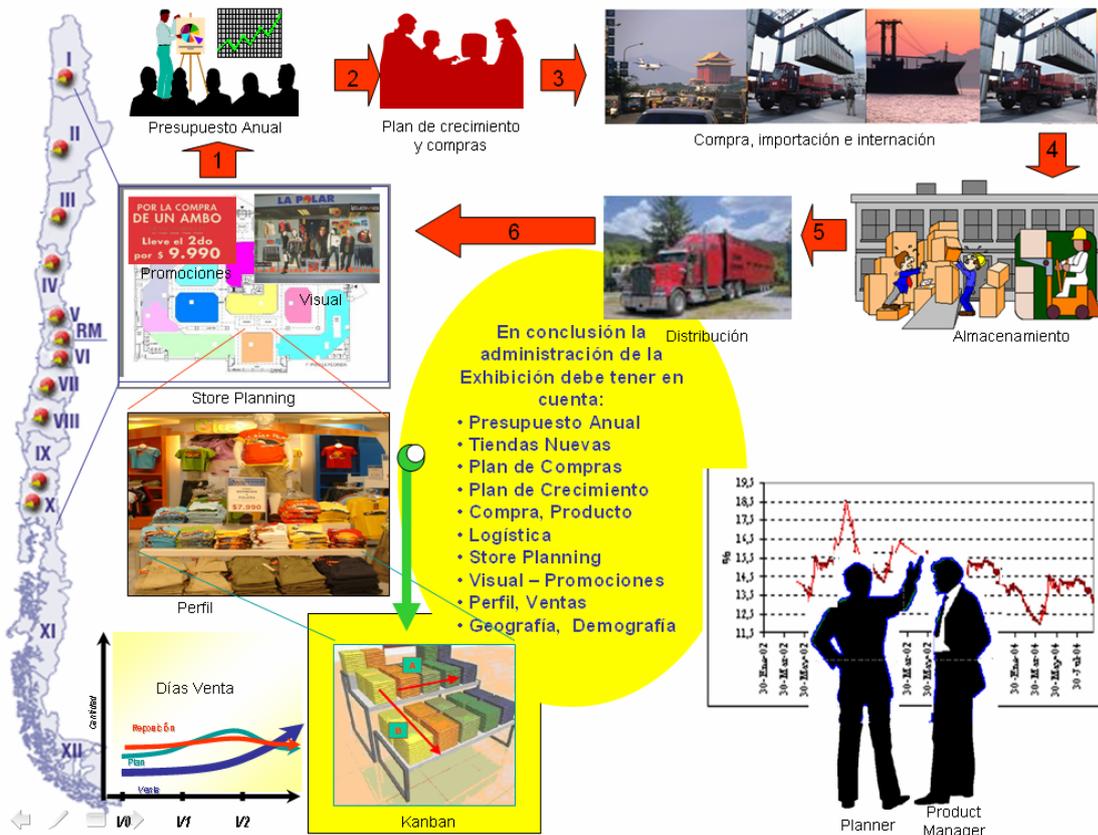


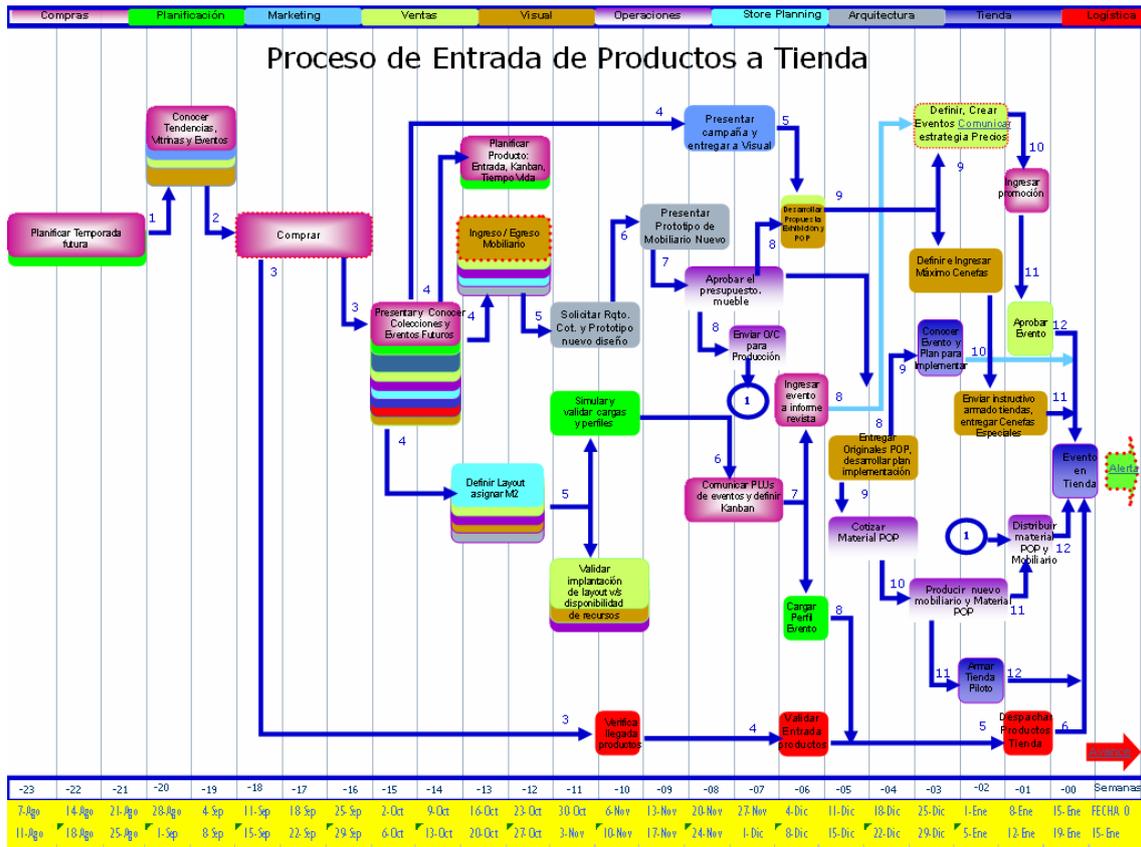
Fig. 4.3 Proceso de Negocios de las grandes tiendas.

Este proceso se desarrolla de la siguiente manera:

1. Análisis del comportamiento de la venta de la temporada equivalente del año anterior y las tendencias futuras
2. Establecimiento del plan de crecimiento para la temporada futura, asignación del presupuesto y autorización de las grandes partidas de compra.
3. Proceso de compra, importación e internación de la mercadería al país.
4. Almacenamiento en bodega central
5. Distribución de la mercadería a las tiendas.
6. Asignación de los espacios de exhibición en tienda, preparación de los puntos de ventas, Análisis y seguimiento a la Venta.

Como podemos apreciar en la siguiente figura, para que un artículo ingrese a tienda, desde que se planifica su compra, el proceso toma 23 semanas aproximadamente, lo cual dificulta realizar lo que indica

la teoría, es decir, hacer compras intermedias, o compras frecuentes, porque es más eficiente.



10 Fig. 4.4 Proceso de Entrada de Productos a Tienda.

En consecuencia, el principal proceso esta centrado en gestionar una adecuada exhibición y fomentar la venta con promociones y eventos. Además se debe fomentar e incentivar la venta con tarjetas de crédito de la misma casa comercial para obtener ingresos adicionales por concepto del negocio financiero, que en algunos casos nuevo alcanza hasta el 80% de las ventas, estos eventos y promociones responden a un proceso que toma aproximadamente una semana en se implementado por la tienda, proceso que se grafica de la siguiente forma:

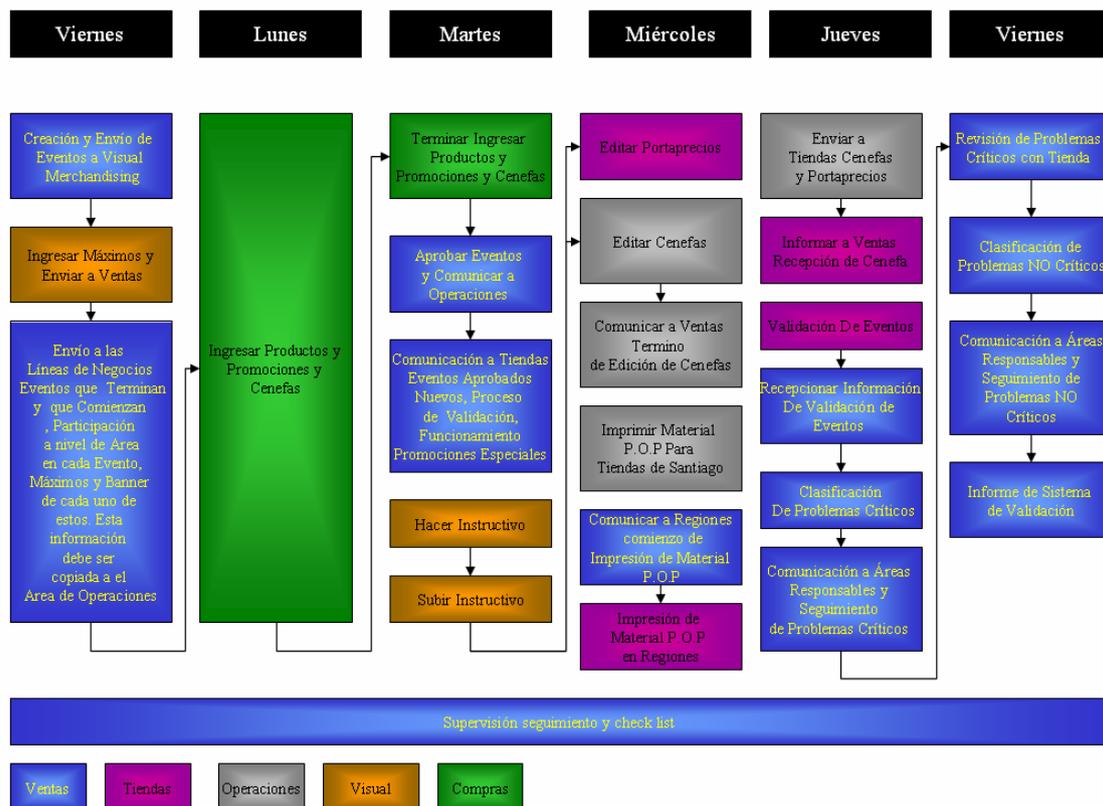


Fig. 4.5 Proceso de Eventos y Promociones.

4.5. Las principales problemáticas de la Venta al Detalle

El análisis del modelo de negocios y de los procesos implantado por las grandes tiendas, que corresponden en general a las mejores prácticas que se dan en la industria de venta al detalle, demandan un continuo desafío de toma de decisiones oportunas e informadas. Sin embargo, si bien la oportunidad es habitual, bien por acción o por reacción, es decir, basta que una de las cuatro líderes de la industria realice un movimiento estratégico para que las otras reaccionen, en este sentido, cada una de ellas esta en permanente vigilancia para mantener o aumentar su participación de mercado.

El gran problema es hacerlo con rentabilidad, es decir, no es suficiente haber obtenido una gran participación de la venta de la industria o mantener cautivos a la mayor cantidad de clientes, lo importante es hacerlo con la mayor rentabilidad posible, y es allí, donde el conocimiento del negocio es vital.

Para aumentar la rentabilidad del negocio se recurre a dos estrategias básicas, la primera es fomentar el negocio financiero, es decir, que el cliente no use el dinero propio, sino que utilice el financiamiento de la tienda, para ello se analiza en el capítulo 5 el caso bajo estudio.

Por otro lado se debe vigilar el Negocio de la Venta propiamente tal, para ello, debe cuidar los tres indicadores básicos: GMROI, GMROS y GMROL, es decir, debe conseguir buenos índices en la gestión del inventario, los puntos de venta y el personal. De estos tres indicadores, el más complejo de administrar por el volumen de datos que debe procesar y lo heterogéneo de las variables que intervienen es el GMROI, fundamentalmente porque las variables que intervienen en las decisiones que afectan a los índices GMROS y GMROL son más evidentes para sus analistas y cuentan con la sensibilidad en terreno de los Gerentes de Tienda y la fuerza de venta. Sin embargo, en la gestión de inventario existe un proceso "oculto" que toma 23 semanas y parcialmente se evidencia su resultado en 6 semanas que toma el ciclo de vida de los productos que más retorno otorgan a la compañías como son los productos de vestir importados y para los cuales no es factible corregir la decisión con recompras o almacenamiento para la temporada siguiente.

En ambos frentes, para el Negocio Financiero y para el Negocio de la Venta al Detalle, se requiere mantener vigente el conocimiento del negocio para tomar decisiones oportunas y eficientes. En la actualidad, los análisis son post-mortem dado la dificultad de tomar decisiones en línea o sobre la marcha. Los analistas se enteran de los comportamientos de compras y de nuevas las tendencias cuando los hechos están consumados.

Habitualmente cuando requieren realizar un estudio de algún comportamiento específico deben recurrir a informática para que les provea una muestra de los datos y con esa muestra realizar sus propios procesos de análisis, utilizan tecnologías de información que les proveen herramientas estadísticas tradicionales y básicas.

Por otra parte, cada línea de negocio posee características propias que son conocidas por cada uno de los analistas quienes en ausencia de información oportuna, toman decisiones basados en la intuición y en la experiencia de las temporadas anteriores. El hecho que cada una de las líneas de negocio posea características propias que varían de acuerdo a una serie de circunstancias, representa un impedimento para que las actuales tecnologías en uso puedan incorporarlos a procesos automatizados. Las capacidades de la infraestructura informática esta dimensionada para soportar el volumen de las transacciones del día a día, en ningún caso para realizar por algún medio conocido gestión sobre las transacciones en curso, todo ello, ha impedido desarrollar una solución que apoye el proceso de toma de decisiones y por lo general la organización dicta políticas y directrices generales y cada analista lo implementa según su intuición y conocimiento de su línea de negocios, en general se trabaja sobre la excepción y sobre montos de venta. No es factible en las

condiciones descritas realizar labores a nivel de ítems o tomas decisiones pro-activas.

Para resolver esta problemática, en el capítulo siguiente se plantea una metodología tendiente a mantener vigente el conocimiento del negocio de manera oportuna, rescatando este conocimiento útil y novedoso desde las bases de datos transaccionales y actualizando el modelo de negocios.

4.6. Resumen del Ámbito de Estudio

En este capítulo se han revisado las características generales de la industria del retail, el detalle que lo sustenta se encuentra en el anexo B. Se han presentado las tendencias generales y se ha descrito el ámbito de estudio, el actual modelo de negocios y su cadena de valor, como también las principales problemáticas que enfrentan los participantes de esta industria.

Capítulo 5

Metodología Propuesta

Este capítulo propone una metodología, que en su estricto significado etimológico significa "una explicación del camino para llegar al objetivo", en esta oportunidad, corresponde a la explicación del camino que disminuye la incertidumbre en las decisiones de carácter estratégico en los ámbitos de la gestión de negocios y de la gestión de riesgo operacional.

La metodología propuesta la denominamos Fuzzy Business Intelligence (FBI), por medio de las técnicas de la computación flexible mantiene vigente el conocimiento del negocio en los procesos de toma de decisiones estratégicas, contribuyendo así, a la competitividad de las empresas que forman parte de industrias dinámicas insertas en macroentornos cambiantes.

Esta propuesta, pretende superar la paradoja que viven las grandes compañías que poseen enormes bases de datos transaccionales, sin embargo, no cuentan con mecanismos apropiados que les permita usar estos datos para resolver las incertidumbres que se presentan al momento de tomar decisiones de carácter estratégico, lo cual, las hace dependientes del buen juicio que aplican los expertos o especialistas que recurren mayoritariamente a la intuición.

Este capítulo describe los procesos y las técnicas que se han considerado apropiadas para la metodología propuesta.

Ejemplos de la aplicación de la metodología se muestran en los Capítulos 6 y 7, el primero esta relacionado con uno de los procesos de gestión de riesgos operacional que realiza el área de control interno, el segundo se desarrolla en el área de negocios aplicado a uno de los procesos de gestión de los resultados del negocio.

La metodología FBI, es una solución integral, dado que cubre los aspectos computacionales e informáticos y los aspectos relacionados con el análisis de los ámbitos donde es aplicada.

Este enfoque permite mantener vigente el conocimiento del negocio, provee en forma oportuna la información útil y novedosa respecto de sus nuevos patrones y tendencias, reduce la presencia de información imperfecta en el proceso de toma de decisiones de carácter estratégico.

Para reducir la incertidumbre y la presencia de información imperfecta, se propone procesar los datos históricos acumulados

provenientes desde las bases transaccionales, transformándolos en información no evidente, útil, interesante y actualizada que muestre los nuevos patrones y tendencias del macro entorno, de la industria y de la empresa.

La metodología FBI es una solución eficiente, porque hace compatible la solución propuesta con los recursos tecnológicos disponibles, esto es, propone reforzar los procesos transaccionales vigentes, mediante la actualización de la lógica de negocios con los resultados provenientes de los procesos que aplican técnicas de la computación flexible.

Para lograr todo lo indicado, la metodología Fuzzy Business Intelligence (FBI), recurre e incorpora conceptos, técnicas y procesos provenientes de distintos ámbitos y disciplinas: desde el ámbito de los negocios, desde las metodologías de la investigación, desde la minería de datos, desde la computación flexible, desde la estadística y desde la computación tradicional, conformando una compleja figura, tal como se grafica a continuación, sin embargo, organiza en ocho etapas que se muestran en la figura 5.2 y cada una de ella esta compuesta por los procesos que se muestran la figura 5.3, lo que otorga coherencia y fluidez.

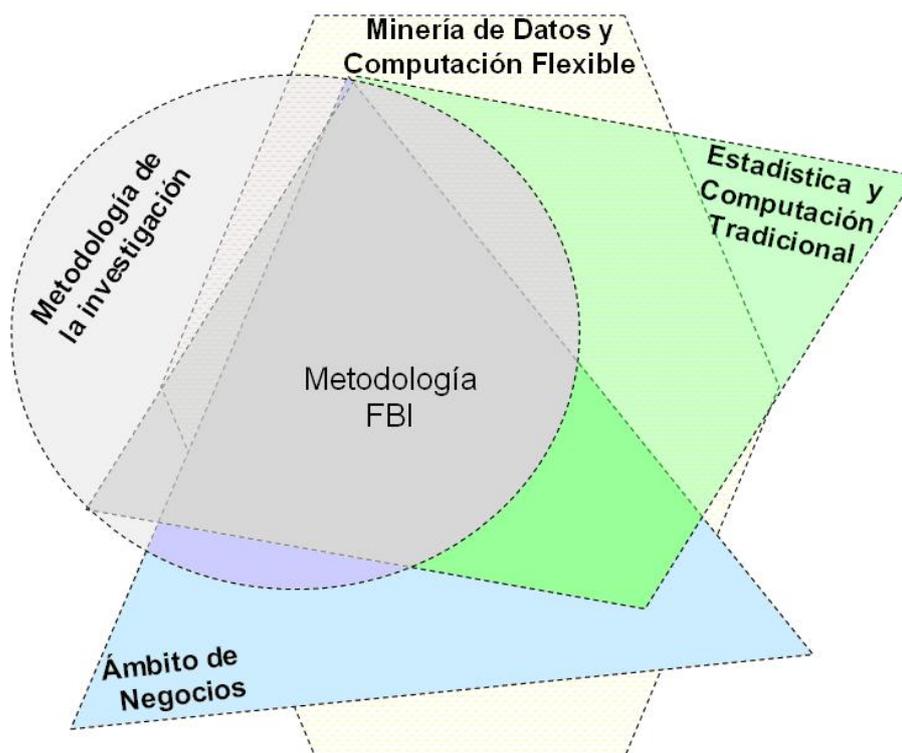


Fig. 5. 1. Ámbitos relacionados con la Metodología Fuzzy Business Intelligence (FBI)

Del **ámbito de los negocios**, FBI incorpora los siguientes procesos:

- Proceso de análisis PEST, plateado por Johnson et al [JG97], el cual permite efectuar un Análisis del macro entorno, para evaluar el impacto o el efecto que ejercen sobre la empresa, el comportamiento y las tendencias presentes y futuras de las variables política, económica, social y tecnológica, este análisis permite construir un escenario presente y un escenario futuro de oportunidades y amenazas, proceso que denominamos: "Construcción de escenarios", sobre los cuales, la empresa define su estrategia competitiva.
- Proceso de análisis de las cinco fuerzas de la Industria, el cual fue denominado por Michael Porter como "Análisis Estructural de los Sectores Industriales", cuya fundamentos y explicación detallada se encuentra en el capítulo 1 de [PM92], lo que permite determinar el nivel de atractividad de la industria, estableciendo las oportunidades y las amenazas que pueden afectar a la empresa en el presente y en el futuro.
- Proceso de análisis de la cadena de valor, modelo también propuesto por Michael Porter y explicado en detalle en el Capítulo 2 "La cadena de valor y la ventaja competitiva" en [PM92a], permite determinar el nivel o potencial de competitividad de la empresa.

Del **ámbito de la metodología de la investigación**, FBI incorpora los conceptos y técnicas relacionadas con la identificación de las variables y de los datos relacionados con el problema bajo estudio, aplicando los contenidos que Fred Kerlinger incluye en el Capítulo 3 "Conceptos, variables y definiciones" en [KF02].

Del **ámbito de la minería de datos y la computación flexible**, FBI incorpora los siguientes procesos y conceptos:

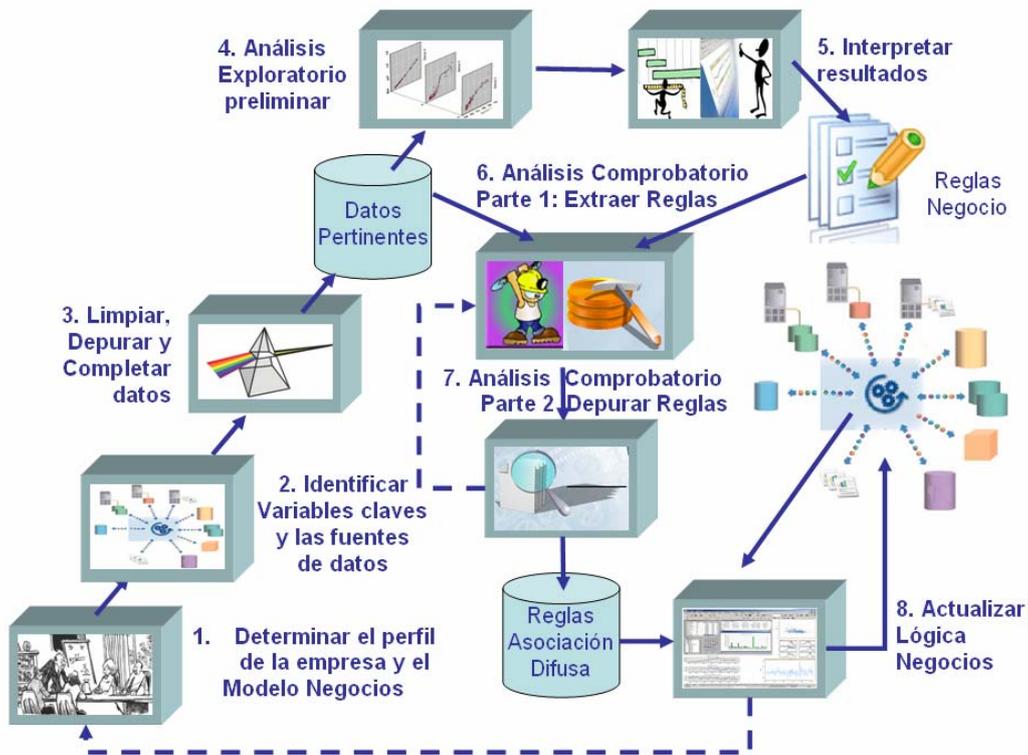
- Aplicación de los conceptos del proceso de descubrimiento de conocimientos en bases de datos, con las técnicas y los modelos de la minería de datos revisados en el capítulo 3 de este documento.
- Aplicación de los procesos de depuración de reglas, usando los métodos objetivos y subjetivos que se analizaron en los apartados 3.4 al 3.6 de este documento.

Del **ámbito de la estadística y la computación tradicional**, FBI incorpora los siguientes procesos y conceptos:

- Los proceso de análisis exploratorios preliminares utilizando técnicas estadísticas mediante herramientas computacionales tradicionales como el SPSS o las planillas de cálculo.
- Los procedimientos recomendados por CMMi¹, para revisar e incorporar en los programas computacionales o aplicativos las nuevas reglas a la lógica de negocios vigente en los sistemas transaccionales.
- Los procesos de control de cambio y aseguramiento de la calidad recomendado por la metodología ITIL², para convertir esta metodología en un proceso recursivo para mantener vigente el conocimiento.

5.1. Descripción de la Metodología FBI

La siguiente figura 5.2., muestra una síntesis de la Metodología FBI:



13 Fig. 5. 2 Síntesis de la Metodología Fuzzy Business Intelligence (FBI)

¹ CMMi : Capability Maturity Model Integration, modelo del Software Engineering Institute, para evaluar la Capacidad y la Madurez de los procesos de generación de software de una organización.

² ITIL: IT Infraestructura Library, es un estándar mundial que reúne las mejores prácticas de gestión de tecnologías de información.

En la siguiente figura 5.3. se muestra en forma resumida los procesos que componen cada una de las ocho etapas de la metodología Fuzzy Business Intelligence:

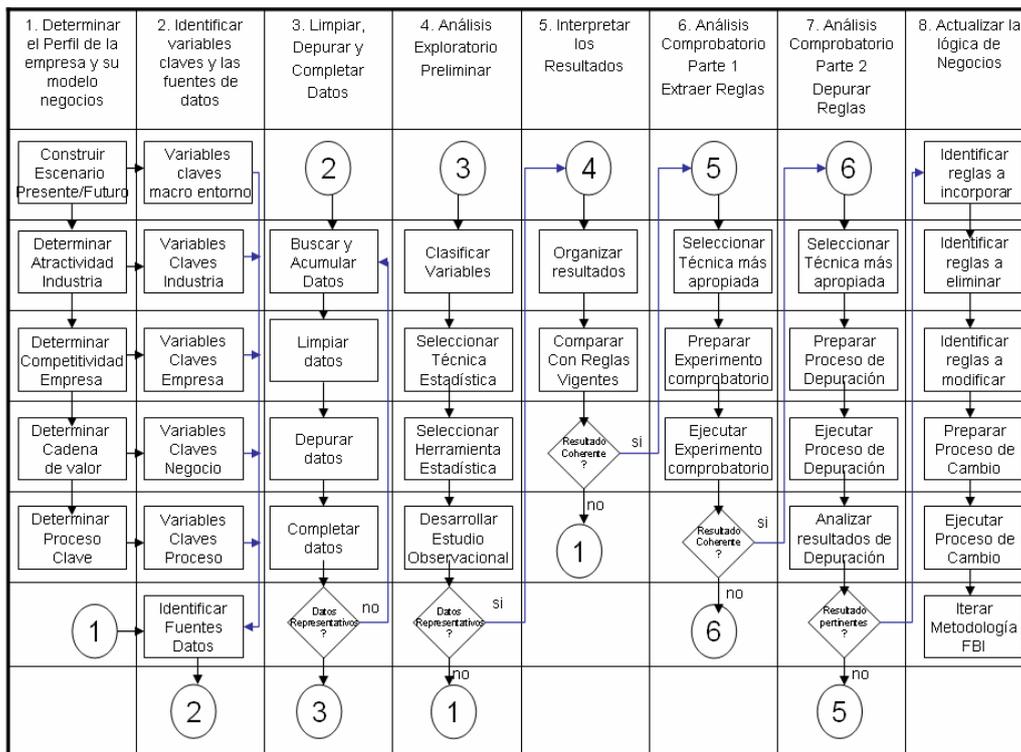


Fig. 5.3. Procesos que componen la Metodología Fuzzy Business Intelligence (FBI)

En los siguientes apartados 5.1.1. al 5.1.8 se explica en detalle cada uno de los procesos que componen cada una de las ocho etapas de la metodología Fuzzy Business Intelligence:

5.1.1. Determinar el Perfil de la Empresa y su Modelo de Negocios

El primer proceso de la metodología propuesta tiene relación con determinar el Perfil de la empresa, esto es, conocer el Nivel de participación y su posicionamiento en el mercado, y también conocer y entender el "Modelo de Negocios".

El Perfil de la Empresa y su Modelo de Negocios, permite comprender como funciona la organización de cara a sus clientes, entendiendo como "Modelo de Negocios", la forma que la organización adopta para generar sus ingresos, es decir, como organiza la empresa su "cadena de valor" para enfrentar las oportunidades y amenazas del macro entorno y de la Industria, teniendo en cuenta sus debilidades y fortalezas.

Análisis Externo:

Para efectuar este Análisis se propone un proceso que va de lo general a lo particular (Top-Down), es decir, desde el macro entorno hacia la empresa.

La siguiente figura 5.4, grafica este proceso de análisis de arriba-abajo, de ella se desprende que: el Macro Entorno es el supra sistema donde esta inserta la empresa que a su vez pertenece y esta inserta en el Sistema Industria (el Sistema Industria lo integran todas aquellas empresas que resuelven la misma necesidad para un segmento objetivo determinado).

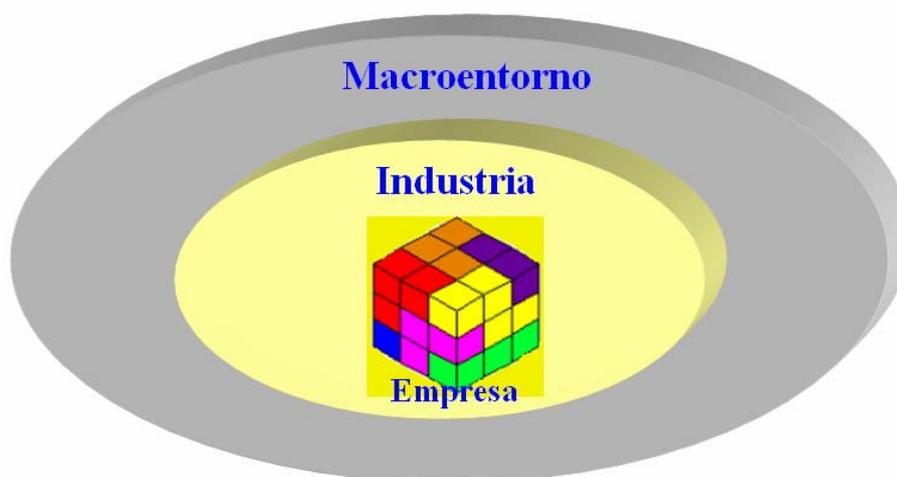


Fig. 5. 4 Sistema: Macroentorno, Industria, Empresa.

Este proceso de análisis, se inicia aplicando el proceso de Análisis PEST explicado en detalle en el Anexo E, el Análisis PEST culmina con la “Construcción de Escenarios”.

La Construcción de Escenarios, permite determinar que tan favorable es el estado actual y futuro del macro entorno para la empresa.

El escenario será favorable, neutro o desfavorable para la empresa dependiendo del estado y de la tendencia de cuatro macrovariables, éstas son: la política, la económica; la social; y la tecnológica.

Cada una de las macrovariables, afecta de manera diferente a los distintos tipos de organizaciones y estas condiciones habitualmente no son constantes, por lo que es necesario, que los escenarios para cada empresa se elaboren para el presente y para el mediano plazo.

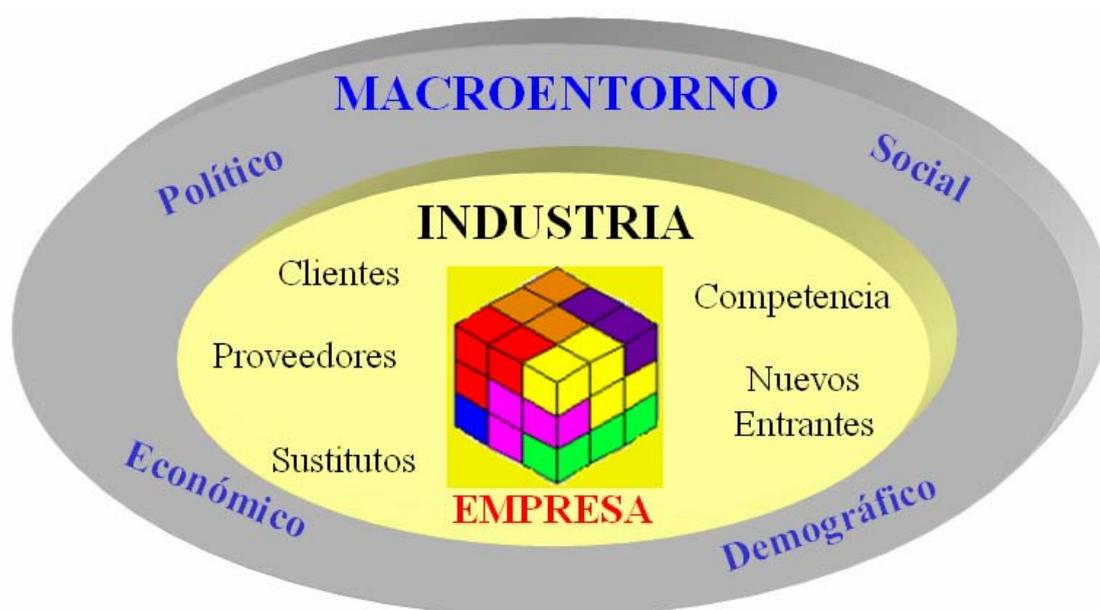
La perspectiva política, se refiere al nivel de exigencias de las regulaciones vigentes y en trámite, las cuales afectarán de diferente manera a la organización, por ejemplo un gobierno proclive a la protección del medio ambiente será menos permisivo con las empresas productivas contaminantes, situación que puede ser permanente en el tiempo.

La perspectiva económica, considera por ejemplo aspectos tales como la paridad cambiaria, si esta es baja y tendiente a la baja afectará positivamente en el presente y en el futuro inmediato al importador y de manera negativa al exportador.

Por su parte, la perspectiva social, tiene relación con las condiciones generales de la población, su cultura, su comportamiento demográfico, por ejemplo, si existe un clima social que aprecie la contribución de la organización a la sociedad y sus prioridades, será favorable para el desarrollo de la organización; si una sociedad que considera prioritaria la educación, o la nutrición será más favorable para una empresa cuyo objetivo sean la prestación de servicios educacionales o la producción de alimentos con ciertas características nutricionales positivas; en el aspecto demográfico, permite conocer y determinar que tan favorable es el comportamiento de las poblaciones, por ejemplo el aumento de la población anciana o la disminución de la población infantil y el impacto positivo o negativo que ello puede tener para el mercado objetivo de la empresa en la actualidad y en el mediano y largo plazo.

Por otro lado, la perspectiva tecnológica, permite conocer y determinar que tan favorable es para la empresa el uso o no de las tecnologías emergentes, analizando que ocurre si se adopta o introduce en la empresa o en el mercado objetivo o en la industria y como ella afectaría en la actualidad y en el mediano plazo a la empresa.

En definitiva los resultados del proceso de Análisis PEST permite la Construcción de Escenarios para estimar cuales son las variables o los criterios provenientes del macro entorno que los expertos o los especialistas consideran al momento de tomar decisiones, en el anexo E, se muestran gráficas y formularios para ilustrar los resultados de estos procesos.



.Fig. 5.5 Componentes de Sistema: Macroentorno, Industria, Empresa.

Una vez conocido y entendido como afecta el macro-entorno a la empresa en el presente y como se estima que éste la afectará en el futuro, se da inicio al proceso de análisis de la atractividad de la industria.

Entendemos por industria al conjunto de empresas que producen el mismo satisfactor, para su análisis, se utiliza el proceso denominado por Porter: "Análisis Estructural de los Sectores Industriales", cuya explicación detallada, tal como se plateo anteriormente, se encuentra en el capítulo 1 de [PM92].

Este proceso de Análisis de las 5 fuerzas de la Industria, es similar al Análisis PEST, en el sentido que realiza un análisis de la situación actual y una estimación de la situación futura de cómo la empresa es afectada por Los Clientes; Los Proveedores; Los Sustitutos; Los competidores; y los Nuevos Entrantes.

Para cada una de estas cinco fuerzas de la industria, se evalúa como afecta y como se estima que afectaran en el futuro a la organización, pudiendo tener un efecto favorable, neutro o desfavorable, para el presente y para el futuro.

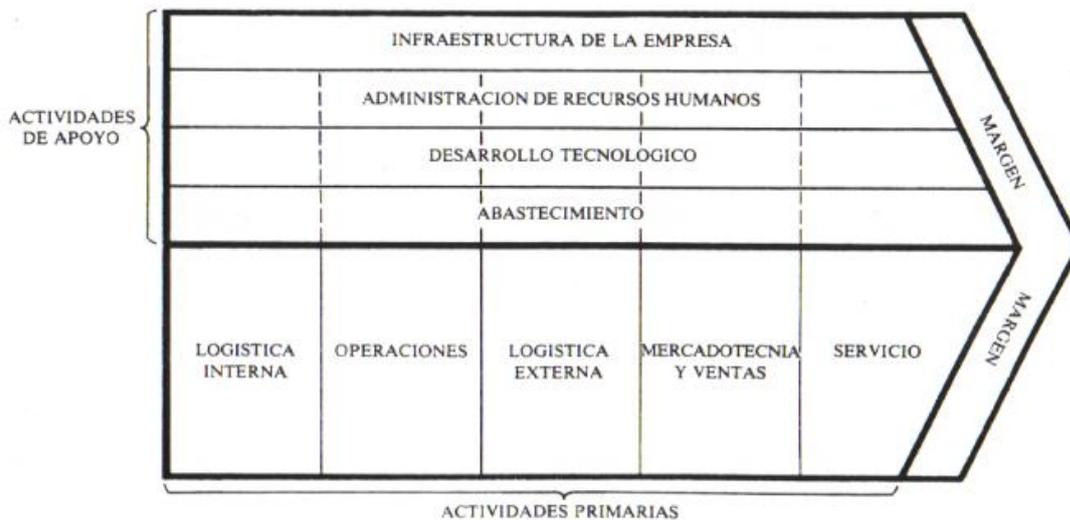
Este proceso permite determinar las oportunidades y las amenazas que se presentan para la empresa. En el Anexo E, se incluyen los formularios para desarrollar este proceso.

Análisis Interno:

Una vez concluido el proceso anterior, denominado Análisis externo, se procede a realizar el denominado Análisis interno, esto es, conocer y entender la cadena de valor de la organización, con el propósito de verificar donde la empresa presenta fortalezas y debilidades.

La forma en que cada empresa organiza sus procesos, es la forma en la cual la empresa pretende crear valor y generar diferenciación ante la competencia para sus clientes, es por ello, la importancia de conocer y entender el funcionamiento de la cadena de valor de la empresa, que en su forma más tradicional Michael Porter la gráfica de la forma que la muestra la figura 5.6, siguiente.

La gráfica anterior, tal como fue mencionado, Porter la explica en detalle en el Capítulo 2 "La cadena de valor y la ventaja competitiva" en [PM92a].



.. Fig. 5. 6 Cadena de Valor. Fuente: Michael Porter[PM92].

En la cadena de valor se aprecia como se distribuyen las actividades de la empresa, las que se organizan básicamente en dos grupos:

Las actividades de apoyo: que corresponde a todas aquellas actividades que contribuyen en forma indirecta a la generación del producto o servicio, muchas de ellas, son susceptibles de externalizar o encargar a terceros, que puedan hacer mejor estas actividades, ya sea, porque la empresa externa es especialista en el rubro o porque tiene capacidad para generar economías de escala, como por ejemplo en la informática; arrendar espacio y servicios de datacenter, en lugar de tener que construir y administrar uno propio, o los servicios generales, tales como: servicios de vigilancia o aseo, servicios especializados como la selección, reclutamiento y administración de personal.

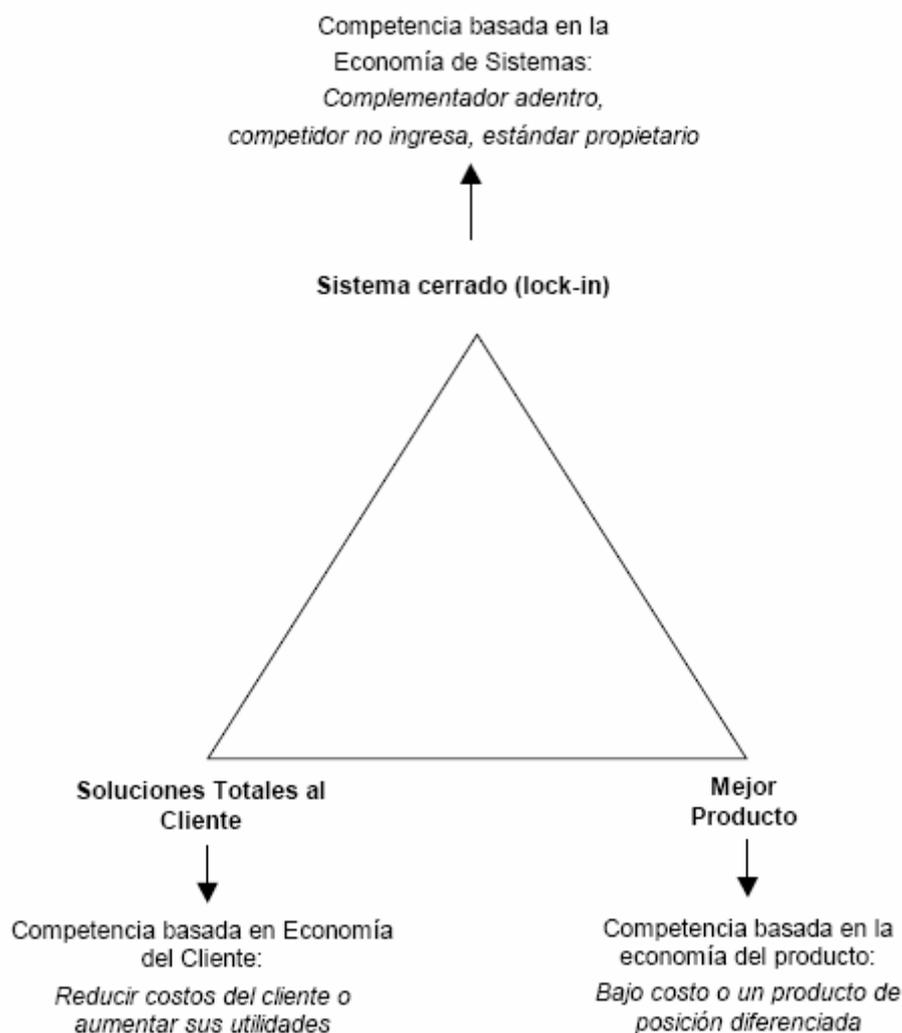
Las actividades primarias: que corresponden a todas aquellas actividades propias del giro de la empresa y relacionadas con la generación del producto o la prestación del servicio de una forma que caracteriza a la empresa y le otorga un sello distintivo frente a sus clientes y una diferenciación frente a sus competidores.

Para concluir este Análisis interno, se debe conocer las definiciones estratégicas de la organización se ha dado en términos de producto, mercado objetivo y tecnología, entendiendo por producto, al resultado que la empresa ofrece al cliente, sea este un bien o un servicio con el valor agregado o sello de la empresa, por mercado objetivo al conjunto de clientes al cual esta dirigido el producto y la tecnología a la forma adoptada por la empresa para producir su producto o servicios.

Estas definiciones, por lo habitual, se encuentran en la declaración de misión, visión, objetivos estratégicos, metas de la organización y las estrategias elegidas para su cumplimiento.

Habitualmente estas definiciones son consecuencia de un posicionamiento estratégico que corresponden a alguno de los vértices descritos por el Modelo Delta de Hax[HA01], que se muestran y explica la figura 5.7.

Este proceso de conocer y estudiar el Perfil de la Empresa y su Modelo de Negocios, aplicando un procedimiento que va desde lo general a lo particular, realizando los Análisis internos y externos, además de conocer y entender la forma que la empresa se ha dado para insertarse en el macroentorno y en la industria, permite interiorizarse del glosario interno facilitando la interpretación de los conceptos y las variables que emplean los analistas o los especialistas y la pertinencia que éstos le otorgan a cada uno de ellos en el proceso de toma de decisiones.



..Fig. 5. 7. Modelo Comercial: Tres Perspectivas Estratégicas. Fuente: Arnoldo Hax[HA01].

5.1.2. Identificar las Variables Claves y las fuentes de datos.

Efectuado los procesos de la etapa anterior, se esta en condiciones de desarrollar lo propuesto por Kerlinger en [KF98], quien en el Capitulo 1, de su obra señala con relación a la búsqueda de conocimiento para el estudio de problemas complejos: "Para entender cualquier actividad humana compleja, es necesario dominar su lenguaje y acercarse a aquellos que la realizan".

Para dominar el lenguaje según señala el mismo Kerlinger en el Capitulo 3 "Conceptos, variables y definiciones", es necesario conocer determinar los conceptos y las variables utilizadas, como también sus definiciones o interpretaciones, de esta forma se inicia la etapa del proceso de obtención de conocimiento cierto, válido, confiable y preciso, que tiene como punto central de todo trabajo de búsqueda de conocimiento: saber **lo que** se busca, bajo que **perspectiva** se va buscar, saber **para que** se quiere encontrar.

Atendiendo al contenido del texto precitado de Kerlinger, recomendamos iniciar este proceso de identificación de datos y sus fuentes, con una entrevista a los especialistas para determinar y conocer los procesos que realizan y cuales son los paradigmas o las afirmaciones en las cuales basan sus decisiones.

En forma habitual los especialistas expresan sus afirmaciones en un formato similar a las reglas de asociación explicadas en el apartado 3.1, y como tal, son afirmaciones o aseveración en forma de conjetura.

Estas reglas de negocios, son proposiciones tentativas sobre la relación entre dos o más variables, como por ejemplo: "Si tal cosa ocurre, entonces se obtendrá tal y cual resultado".

Estas aseveraciones habitualmente planteadas en forma de oraciones declarativas, relacionan variables con variables, ya sea en forma general o específica.

La aseveración conllevan contrastes para probar las relaciones establecidas, esto significa que la aseveración contienen dos o más variables medibles o potencialmente medibles, y que especifican al mismo tiempo de que manera se relacionan dichas variables.

Se debe tener presente que las variables que forman parte de estas reglas de negocios son, por lo general, conceptos o propiedades propios del ámbito de estudio.

También se debe tener en cuenta que una variable es un constructo o una propiedad que adquiere distintos valores, en otros términos una variable es un símbolo al cual se asignan numerales o valores.

La taxonomía de variables según Kerlinger es la siguiente:

Variables Independientes y dependientes: la variable independiente es la supuesta causa de la variable dependientes y ésta el supuesto efecto. La variable independiente es el antecedente y la dependiente es el consecuente. Cuando se dice Si A entonces B, se tienen una unión condicional de una variable independiente (A) y una dependiente (B). En términos de experimentos la variable independiente es la variable manipulada por el experimentador.

Variables Activas y Atributivas: otra clasificación útil en una investigación son las variables manipuladas llamadas activas y las variables medidas, llamadas atributivas. Ejemplos de variables atributivas, es decir, aquellas que no son manipuladas, son sexo, inteligencia, aptitud, estatus socioeconómico, etc.

Variables Continuas y Categóricas: una variable contigua es capaz de tomar un conjunto ordenado de valores dentro de cierto rango. Las variables categóricas pertenecen a una clase de medidas conocida como nominales, todos los miembros perteneciente a la categoría posee la misma característica que define el subconjunto.

Para ejemplificar este proceso, presentamos alguno de las reglas de negocios planteadas por los especialistas durante las entrevistas de los casos bajo estudio en los capítulos 6 y 7:

“Si la compra se realizo con una tarjeta de reciente apertura y pertenece a una persona joven de sexo femenino entonces es muy probable que la impugnación sea efectivamente un fraude”

“Si una tienda grande en un mall exhibe todas la tallas y todos los colores de una prenda entonces la prenda será ganadora”

En ambas reglas de negocios se aprecian conceptos propios de cada área de negocios y dominios poco frecuentes para algunos de ellos, situación que se analiza en detalle en cada uno de los casos ejemplo.

Determinado el conjunto de afirmaciones o aseveraciones o paradigmas, en las cuales los analistas, especialistas o expertos, basan sus decisiones, y habiendo identificado los conceptos o

variables que intervienen y como ellas se relacionan, se esta en condiciones de identificar los datos y sus fuentes.

Una vez conocidas las variables, se necesitan ubicar los datos asociados a su comportamiento histórico, para ello se requiere entrevistar a los representantes del área de la computación en aquellas organizaciones donde el nivel de madurez de la informática de gestión es menor o igual de tres y al área de la informática en los casos que la madurez de informática de gestión sea superior al nivel tres.

En esta entrevista es conveniente determinar el nivel de madurez de la informática de gestión (descritos en el apartado 2.9.2), dado que este nivel de madurez tiene una relación directa en el nivel de concentración y en ocasiones con la calidad de los datos a obtener.

El nivel de dispersión y el nivel de calidad de las fuentes de datos, tiene relación con el nivel de madurez de informática de gestión que la organización tiene, a menor nivel de madurez mayor dispersión y menor calidad de los datos históricos.

En general los datos históricos se encuentran dispersos por la organización, en la mayor parte de las empresas se debe recurrir a las diferentes tecnologías de tratamiento y almacenamiento de información que la organización utiliza en sus sistemas transaccionales que apoyan la operación diaria.

Si la empresa no se encuentra en un nivel avanzado de madurez de gestión informática, esto es, inferior al nivel 5, las fuentes de datos son heterogéneas e incompletas, lo cual, da origen al paso siguiente de limpieza, depuración y completitud de los datos obtenidos.

5.1.3. Limpiar, depurar y completar los datos.

Conocidas las fuentes de datos, se procede a extraerlos, organizarlos de manera preliminar en un repositorio con los datos históricos, representativos y pertinentes al proceso y al período bajo estudio.

En esta etapa aparecen las imperfecciones de los datos históricos, éstas imperfecciones o deficiencias se dan por lo general por datos erróneos, incompletos, imprecisos, inexactos, dado que lo habitual es, que las fuentes de datos evolucionan según sean las necesidades del negocio y las restricciones impuestas por los recursos computacionales e informáticos disponibles en cada una de las fracciones de tiempo que componen el periodo bajo estudio. Frente a este escenario es necesario identificar estas imperfecciones o deficiencias, dando origen a la tarea de extraer, limpiar, depurar y

completar los datos, para ello se recurre a las diversas técnicas de tratamiento de datos con el propósito de contar con datos homogéneos para el período analizado, esto va a significar en muchos casos a convertir, calcular o generar datos a partir de otros, usando desde planillas de cálculo hasta sofisticados algoritmos incrustados en los ambientes de bases de datos, en esta etapa el proceso de minería de datos debe incorporar un proceso de restauración de los datos faltantes, perdidos o incompletos.

5.1.4. Análisis exploratorio preliminar.

Una vez que se cuenta con los datos pertinentes, es decir, se ha obtenido un repositorio o datawarehouse o datamart, se inicia el proceso de análisis exploratorio preliminar, para ello se recomienda recurrir a herramientas tradicionales de análisis estadístico de datos como por ejemplo SPSS o a algunas de las funcionalidades de las planillas de cálculo como por ejemplo las de Microsoft Excel u otro similar.

Este análisis exploratorio preliminar permite validar la consistencia y pertinencia de los datos obtenidos y generados como resultados de los procesos anteriores.

Este análisis preliminar otorga una visión global del comportamiento y dominio de las variables bajo estudio, permite comprender mejor la problemática y obtener los elementos que permiten realizar una interpretación preliminar de los resultados.

5.1.5. Interpretación de resultados.

Concluido el análisis exploratorio preliminar se procede a la interpretación de resultados, para ello se trabaja en conjunto con los especialistas y se analiza el resultado del proceso anterior, comparando estos resultados con el conjunto de reglas de negocios que los sistemas transaccionales tienen definidas en sus programas computacionales y los paradigmas que aplican o las interpretaciones que estos especialistas indican como pertinentes.

Eventualmente en esta etapa, como producto de la aplicación de paradigmas o interpretaciones puntuales, aparecen consideraciones adicionales que los expertos o especialistas toman en cuenta y que no necesariamente están sistematizados en la lógica de negocios residente en los sistemas transaccionales, y dan origen a nuevas variables cuyos datos son obtenidos desde otras instancias, por ejemplo, planes de negocios o planes comerciales cuyos parámetros no forman parte del modelo de datos transaccional, completando así, el conjunto de datos necesarios para este proceso, en la eventualidad

que esta situación se presente, se debe volver a iterar el proceso descritos en el apartado 5.1.2.

De esta manera, esta instancia, además entrega la oportunidad de verificar la vigencia y completitud de las reglas del negocio con los actuales criterios y/o las políticas comerciales que se ha impuesto la organización, originan modificaciones, actualizaciones y eliminaciones de reglas de negocios de los sistemas transaccionales y/o la modificación de los paradigmas o afirmaciones que los expertos o especialistas aplican.

Obtenido de esta forma el nuevo repositorio depurado y habiendo verificado la pertinencia, completitud y vigencia de las variables, se procede con el siguiente proceso, correspondiente al análisis comprobatorio.

5.1.6. Análisis Comprobatorio (Parte 1: Extraer Reglas).

Hasta ahora, los procesos de la metodología propuesta, han utilizado herramientas convencionales o tradicionales. Para esta etapa, la metodología propone, realizar en primer lugar, un Análisis de los métodos, las mejores prácticas y las herramientas de última generación disponibles, para determinar cual de ellas, es la más apropiadas o adecuadas para buscar el conocimiento no evidente en los datos históricos acumulados.

Las características propias de cada situación bajo estudio determinan los métodos o técnicas más favorables o apropiadas para desarrollar el análisis comprobatorio, en el anexo A, se presenta un inventario no exhaustivo de ellas, como ejemplo, podemos mencionar, el caso de Parisi et al [PA07], donde se evalúan varias de las técnicas avanzadas, en particular, las más adecuadas para la predicción de la variación del Ishre, analizando la eficacia de los modelos multivariados dinámicos, contruidos a partir de recursividad, Rolling, algoritmos genéticos, redes neuronales, fractales y lógica difusa, este caso, sirve para ejemplificar que un caso puede aplicar más de una técnica y que sus resultados pueden arrojar resultados significativamente similares o distintos según sea el propósito de su aplicación y las características de las variables bajo estudio.

En los dos casos de estudio presentes en este trabajo, se revisan y analizan los modelos multivariados para determinar la certeza de las reglas de negocios o afirmaciones que aplican los especialistas para predecir el comportamiento o conducta de las ventas y de las impugnationes o los desconocimientos de compra desde la

perspectiva de la gestión de negocios y de la gestión del control interno, aplicando las reglas de asociación difusa.

La elección de las reglas de asociación difusa, como técnica de minería de datos, en ambos casos, obedece fundamentalmente al hecho que los elementos claves del pensamiento que aplican los analistas o expertos no son numéricos, sino que son ideas con cierto grado de vaguedad, donde los elementos pasan de un conjunto a otro de manera suave y flexible, demandando una herramienta o técnica que sea atractiva para manejar la incertidumbre.

En estos casos, las decisiones estratégicas bajo estudio, son ejemplos paradigmáticos de dominio incierto, aunque todas las fuentes de incertidumbre se pueden dar, sobre todo con el uso del lenguaje natural donde la información implícita, la multiplicidad, la ambigüedad y la imprecisión hacen imprescindible el tratamiento de la incertidumbre.

Para la aplicación de la técnica de minería de datos, seleccionada, es decir, la extracción de reglas de asociación difusas, se requiere en primer lugar compatibilizar los datos con las variables y conceptos, para ello se recurre a los conceptos y técnicas que se explican en el Anexo C, lo que permite determinar las etiquetas lingüísticas que participaran en este proceso y sus rangos o dominios, que corresponden al intervalo sobre el cual se define la función de pertenencia.

Una de las principales utilidades de las etiquetas lingüísticas es que pueden incluir muchos valores posibles, lo que ayuda a caracterizar fenómenos complejos de definir, como es en los casos que se presentan en el capítulo 6 y 7.

Definidas las etiquetas lingüísticas, lo que permite hacer imprecisas las variables difusas, se aplican los algoritmos de extracción de reglas de asociación, en este caso se recurre al software FuzzyQuery 2+ de Serrano et al [SJ03] y al software CBA-Aias de Lui [LB99].

Aplicada la técnica de minería de datos seleccionados, para extraer los patrones de comportamiento y las tendencias que los datos acumulados no muestran en forma explícita, se extraen las reglas de asociación.

Una de las características de la técnica seleccionada, es la gran cantidad de reglas de asociación que se generan producto del cruce de todas sus variables, para hacer más eficiente el trabajo de depuración de las mismas con los especialistas, se requiere aplicar criterios que permitan depurarlas en forma sistemática, de manera de reducir su volumen y que efectivamente conduzcan a obtener reglas

útiles y novedosas, para ello se recurre a las técnicas cuantitativas y cualitativas, procesos que se aplica en la segunda parte del análisis comprobatorio y que se revisa en el apartado siguiente.

5.1.7. Análisis Comprobatorio (Parte 2: Depurar Reglas).

Concluida la primera parte del Análisis Comprobatorio, se inicia un proceso recursivo para depurar las reglas de asociación difusa extraídas.

Durante esta etapa, se desarrollan dos grandes procesos, el primero corresponde al de selección de reglas de asociación difusa fuertes y muy fuertes.

Las reglas fuertes y muy fuertes se obtienen aplicando los algoritmos propuesto por Berzal et al [BF02], basadas en las condiciones de soporte y confianza, que se traduce en, la aplicación de la valoración objetiva a las reglas de asociación, este proceso esta descrito en el apartado 3.5.

Al subconjunto resultante de reglas fuertes y muy fuertes, se aplica el proceso de valoración subjetiva para reglas de asociación difusa definido por Liu[LB93] y Sahar[SS00].

El proceso de aplicación de medidas de interés subjetivas de Liu, consiste en aplicar las jerarquías de similitud:

- a) Jerarquía según la similitud de los ítems del antecedente de las reglas obtenidas en el proceso de minería de datos con los ítems del antecedente de las reglas definidas por el usuario
- b) Jerarquía según similitud de los ítems del consecuente de las reglas obtenidas en el proceso de minería de datos con los ítems del consecuente de las reglas definidas por el usuario
- c) Jerarquía según similitud de los ítems de ambos lados de las reglas obtenidas en el proceso de minería de datos con los ítems de ambos lados de las reglas definidas por el usuario.
- d) Jerarquía según similitud de conformidad de las reglas obtenidas en el proceso de minería de datos con las reglas definidas por el usuario.

Para aplicar los criterios de jerarquías de similitud se puede recurrir a procesos manuales o a procesos automatizados provistos por Liu en sus aplicativos computacionales CBA-Aias.

El análisis de los resultados de la aplicación de estas medidas de similitud permite a los especialistas clasificar las reglas de asociación seleccionadas en alguna de las siguientes categorías:

No verdaderas y no interesantes

Verdaderas y no Interesantes

No verdaderas e interesantes

Verdaderas e interesante.

Naturalmente las pertenecientes a la última de estas categorías serán la que los especialistas evaluarán para ser consideradas en el proceso de actualización de la lógica de negocios que se revisa en el apartado siguiente.

De este proceso también, los especialistas infieren cuáles son las reglas del negocio ausentes o con baja presencia y que eventualmente estarían obsoletas, lo que les permite determinar el nuevo conjunto de reglas de negocios vigentes.

Como resultado de este proceso se deben generar las instrucciones al área informática para agregar, modificar y/o eliminar las reglas de negocios dando origen a la última etapa de la metodología.

5.1.8. Actualización de la Lógica de Negocios.

Para que esta Metodología tenga un efecto real en la competitividad de la empresa, sus resultados se deben incorporar a los procesos transaccionales tradicionales.

Este proceso de validar y mantener la vigencia de las reglas de negocios, requiere incorporar los resultados obtenidos de las etapas anteriores aplicando los procedimientos de modificación de los sistemas transaccionales siguiendo los estándares y/o normas u otras que la empresa se ha dado para el aseguramiento de la calidad de este proceso, como por ejemplo CMMi y/o ITIL.

5.2. Resumen de la Metodología Propuesta

Este capítulo ha presentado la metodología Fuzzy Business Intelligence, revisa en forma detallada los modelos, las técnicas y los procedimientos que fueron considerados pertinentes para reducir la incertidumbre en la toma de decisiones estratégicas y mantener vigente el conocimiento del negocio, fomentando de esta forma la competitividad de la empresa.

Capítulo 6

Proceso de Gestión de Riesgo

Caso: Detección y Prevención de Fraude en tarjetas con crédito de grandes tiendas.

A continuación se presenta la primera aplicación de la metodología Fuzzy Business Intelligence, que permite mantener vigente el conocimiento del negocio para la toma de decisiones estratégica en el proceso de gestión de riesgos y contribuir a la competitividad de una empresa que forma parte de una de las industrias más dinámicas del cono sur de América y esta inserta en un macroentorno cambiante inmerso en la globalización.

El caso bajo estudio, corresponde al área de Gestión de Riesgo encargada de realizar los Procesos de Control Interno. El proceso específico corresponde al de Detección y Prevención de Fraude con tarjetas de crédito.

Las motivaciones para aplicar la metodología Fuzzy Business Intelligence son las siguientes:

Ausencia de un sistema en línea para detectar y prevenir la actividad fraudulenta, el actual proceso de “detección” de fraude, tiene relación con la verificación de la vigencia de la tarjeta de crédito y su capacidad de endeudamiento y no con una labor pro-activa que prevenga efectivamente la actividad fraudulenta, la mayor parte del análisis de riesgo de fraude se relaciona con las impugnaciones, las cuales corresponden a hechos consumados los cuales según sean las características del cliente pasan a ser de cargo de la empresa bajo estudio, lo que naturalmente encarece los costos de operación y perjudica a todos los involucrados en el negocio.

Toma de decisiones descentralizada para analizar y enfrentar las acciones fraudulentas, los acciones para detectar acciones fraudulentas están descentralizadas, es decir, una vez conocida la impugnación, la empresa bajo estudio encarga a sus analistas de riesgo que resuelvan las acciones fraudulentas de las transacciones relacionados con ventas efectuadas en sus locales de venta, las transacciones efectuadas en las instalaciones de sus socios comerciales deben de ser resueltas por cada uno de ellos y con sus respectivos recursos.

Tendencia al uso de herramientas de software de uso tradicional, la mayor parte de las implementaciones de herramientas de software esta formada por modelos matemáticos y herramientas de análisis estadístico de datos, como por ejemplo SPSS, Intelligent Mining, Model One, los cuales se utilizan para clasificar el riesgo, todo ello con resultados poco intuitivos como son los puntajes o scoring.

Rápido incremento de la actividad fraudulenta, como lo muestra el análisis estadístico preliminar, existe un importante incremento de la actividad fraudulenta (15% anual), y su mayor incidencia esta en transacciones generadas fuera del ámbito de los puntos de ventas propios de la empresa bajo estudio, tendiendo esta actividad a tener comportamiento como los que señalan los estudios de este fenómeno publicados en los sitios especializados ya referenciados.

Necesidad urgente de cambio a una posición pro-activa, la tasa de crecimiento de la actividad fraudulenta conlleva a un rápido deterioro de la relación con los socios comerciales y clientes, como también de un incremento de costos de operación para todos los involucrados, todo lo cual señala la urgencia de pasar de una política reactiva de análisis post-mortem a una proactiva, es decir, una política real de detección y prevención del fraude con un sistema en línea que resguarde los intereses de todos los involucrados.

6.1. Determinar el Perfil y el Modelo de Negocios de la empresa

El resultado del proceso de identificación del perfil de la empresa y la determinación del modelo de negocio, se muestra en el capítulo 4 "Ámbito de Estudio: Las grandes tiendas", en los apartados 4.1 al 4.6, y la descripción detallada de los principales participantes de la industria bajo estudio se incluye en el Anexo D.

De este proceso se concluye que en este caso, el ambito de estudio corresponde a una de las empresas líderes de la industria del retail latinoamericano y que su modelo de negocios, busca incentivar el uso de la tarjeta corporativa como medio de pago y corresponde a una estrategia de fidelización de clientes.

Esto significa que por cada compra que se realiza con este instrumento de pago, el cliente recibe puntos que posteriormente puede canjear por productos.

Esta tarjeta, además de ser el medio de pago más usado en las tiendas propias, se usa en establecimientos asociados que comercializan productos y servicios complementarios como: venta de comida rápida; combustible; productos de farmacias.

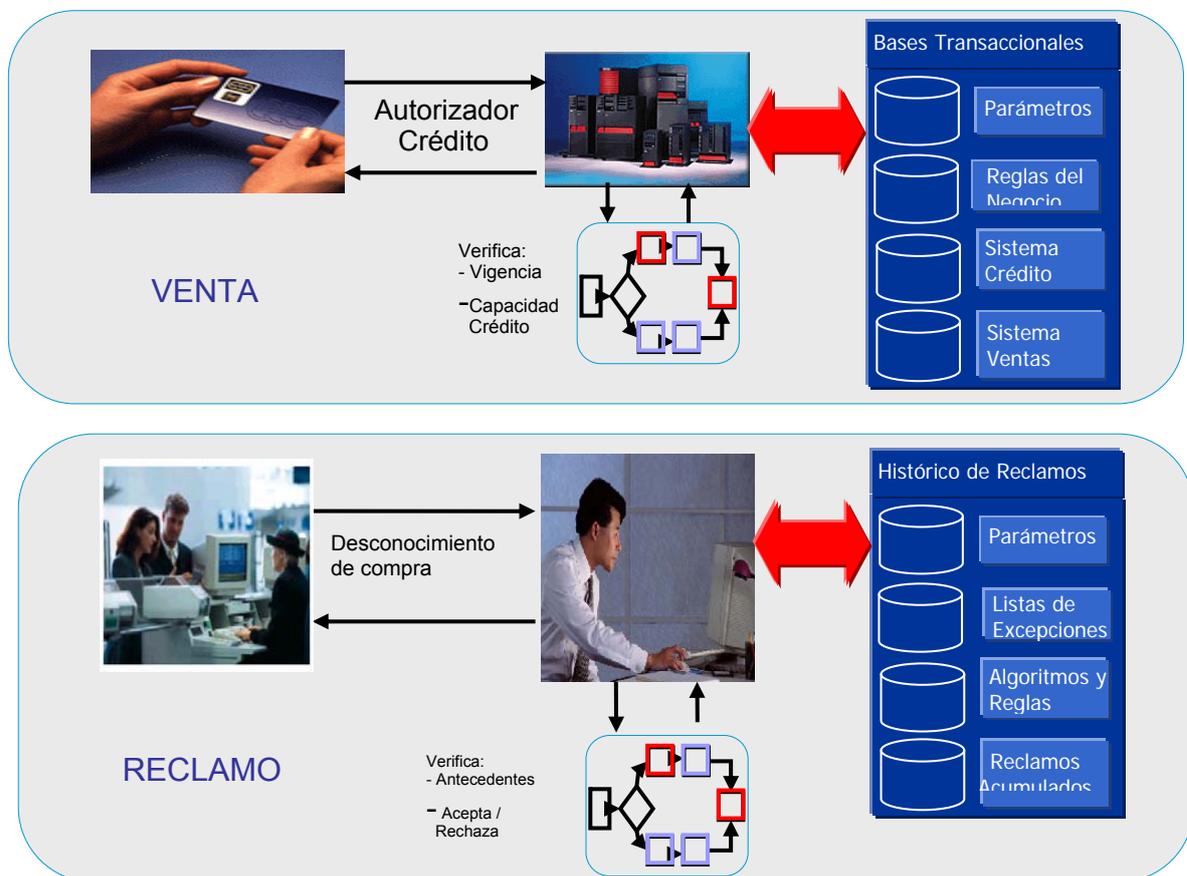
También existen acuerdos que habilitan convenios con empresas de telecomunicaciones para recarga de tarjetas de pre-pago telefónico.

Todo lo anterior, corresponde a una estrategia genérica de habilitar a sus clientes para obtener con un solo medio de pago todos aquellos servicios que necesite y en muchos casos a precios preferenciales.

La empresa bajo estudio, cuenta con alrededor de cinco millones de tarjetas de crédito emitidas, cuyo uso representa en promedio el 30% de las ventas.

En general, a nivel nacional existen en total, alrededor de 16 millones de tarjetas de crédito emitidas por las grandes cadenas comerciales, esto es, siete veces más que la suma de las tarjetas Visa y Mastercard del país [EM04].

En términos de uso, en el país, este tipo de tarjetas supera en promedio el 65% de las ventas, representando casi el 20% de la deuda de consumo del país, sabiendo que existen más de 11.000 establecimientos, no emisores de este tipo de tarjetas que las aceptan [EM04].



.. Fig. 6. 1 Actual proceso de tratamiento del desconocimiento de compra. Fuente: Adaptado del software Monitor

La figura 6.1, muestra en forma grafica el modelo de operación vigente al momento de efectuar la investigación, en esta gráfica, se verifica que, para que la transacción con tarjeta de crédito sea aceptada sólo se requiere que la tarjeta este vigente y cumpla con las reglas de negocio básicas, esto es, que tenga saldo disponible para cubrir el cargo.

También se aprecia, que el proceso de verificación de una situación anormal, es a pedido del cliente, quien hace presente una situación que lo afecta y que se transforma finalmente en un desconocimiento de compra.

Frente a este desconocimiento de compra se activa un procedimiento de verificación y posteriormente el cliente es notificado del resultado de su solicitud.

El uso intensivo de tarjetas de crédito como medios de pago y la vulnerabilidad que estas presentan al no existir medidas pro-activas efectivas para prevenir el uso indebido, requiere que se optimice los métodos y procedimientos que validen el correcto uso, prevengan o detecten en forma oportuna situaciones fraudulentas.

El actual proceso de administración del fraude de la empresa bajo estudio, es posterior a su ocurrencia. Cuando se produce un desconocimiento de compra de una transacción hecha con tarjeta de crédito, la investigación se apoya recopilando datos de varios sistemas que aportan antecedentes del clientes y lo clasifica en distintas categorías.

Los procedimientos usados en este proceso de administración del desconocimiento de compra (eventual fraude), se aplican una vez que el cliente presenta su reclamación en el servicio de post-venta. En estos procesos se combinan la experiencia de los expertos en desconocimiento de compra, con procesos manuales y procesos automatizados, que son combinados con sistemas de información convencionales que apoyan su actividad.

6.2. Identificar las Variables Claves y las Fuentes de Datos

Conocido el perfil de la empresa bajo estudio, su modelo de negocios y el proceso que realizan los especialistas, se esta en condiciones de iniciar el proceso siguiente que permitirá identificar las variables y los conceptos usados por los especialistas en Gestión de Riego y Control Interno, teniendo presente que en este proceso se deben, por lo general, obtener los criterios y paradigmas de negocios que manejan los especialistas y que se presentan en formato similar al de reglas de asociación.

El apoyo tecnológico al que recurren los especialistas en este proceso, corresponde al conjunto de tecnologías de información y plataformas tecnológicas que apoyan la operación y son usados para retroalimentar la toma de decisiones, no cuentan con sistemas de

soporte pro-activos que permitan prevenir un desconocimiento de compra.

Estos sistemas de apoyo indirecto, en su gran mayoría son, sistemas tradicionales cuya función están relacionadas con las operaciones diarias, propias del giro de la empresa, complementados con un conjunto de procedimientos manuales que aplican los expertos.

En los procedimientos manuales que realizan, utilizan algunas herramientas tecnológicas para apoyar las labores de toma de decisiones.

El proceso que realizan los convierte en intermediarios entre los datos provenientes de la operación, el apoyo tecnológico ad-hoc y los ejecutivos de alto nivel que demandan conocimiento para la toma de decisiones estratégicas.

La empresa bajo estudio cuenta con distintos modelos de tratamiento y representación de la información desde archivos planos que utilizan lenguajes procedurales como COBOL, pasando por bases de datos relacionales de última generación como ORACLE, además de utilizar algunos modelos matemáticos y algunas herramientas y productos para tratamiento estadístico de datos como por ejemplo SPSS, Intelligent Mining, Model One, entre otros.

De esta manera, para cada situación o necesidad se cuenta con un conjunto de sistemas tradicionales, procedimientos manuales y herramientas de software utilizados por expertos que usadas en conjunto resuelven puntualmente la necesidad de información como lo representa la figura siguiente.

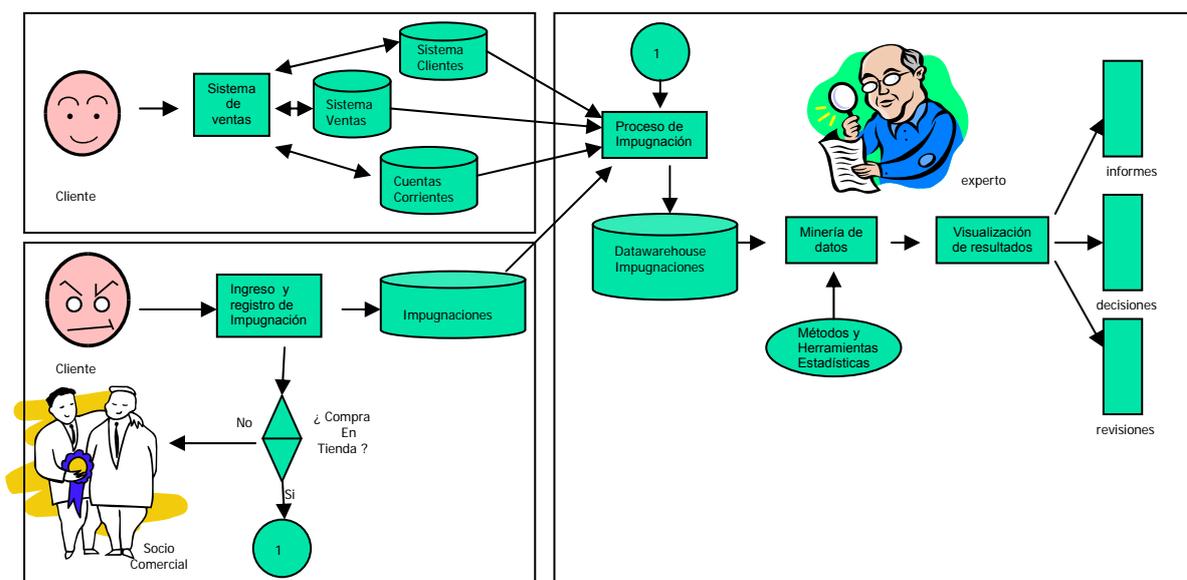


Fig. 6. 2 Proceso de tratamiento de impugnación o desconocimiento de compra.

En resumen, la descripción del proceso anterior nos señala, que existe un proceso orientado a la administración de un problema relacionado con el control de riesgo operacional, sin existir un procedimiento o proceso durante la operación normal del cliente, esto significa que no existe sistema para prevención y detección de fraude. Lo que existe son procedimientos que se activan para atender un reclamo y dependiendo de las características de la transacción, corresponde recurrir a uno u otro proceso que son realizados con posterioridad a la notificación por parte del cliente al que se ha cargado en su cuenta una transacción que él desconoce, esta acción recibe el nombre de "Impugnación".

Para decidir si un desconocimiento de compra se trata efectivamente de un fraude, se inicia un proceso que les permite comprobar la existencia o no del fraude aplicando paradigmas o conceptos como los siguientes:

En relación al dueño de la tarjeta: Los fraudes habitualmente afectan a mujeres jóvenes que usan tarjetas de créditos emitidas recientemente.

Con relación a la compra o transacción: los principales productos involucrados en fraude son productos de fácil reducción a dinero o corresponden a cierto tipo de servicios como carga de combustible, compra en farmacias, consumo de comida rápida o recarga telefónica.

Otros antecedentes que les aportan buenos indicios que se encuentran frente a una situación fraudulenta, es el cambio de comportamiento de compra del cliente: entre las situaciones que les permiten identificar estas condiciones están:

- Cambio de lugar de compra habitual.
- Compras de productos que no corresponden a su perfil de cliente
- Compras en cantidades distintas a la habitual.
- Aumento en la frecuencia de compra de un mismo producto.
- Compras en distintos locales en un corto plazo.
- Compras sin abono.

Los antecedentes recopilados permiten determinar que se requieren datos de Clientes, Productos, Ventas.

Alguno de los procesos que mantienen vigente los datos de interés relacionados con el cliente, y que se realizan en forma periódica en un proceso nocturno, son los correspondientes a la actualización de sus antecedentes personales y comerciales. El proceso de actualización de su saldo en cuenta corriente. El proceso de

actualización del estado de la cuenta y de otros datos comerciales provenientes de fuentes externas como deuda en el sistema financiero, que permiten actualizar su clasificación de riesgo y, comportamiento de pago (scoring).

Otros procesos a los que recurren los especialistas en administración de los riesgos de las operaciones de crédito, son los procesos que manejan los datos de las transacciones que genera un cliente al realizar una compra, en ese momento, se activa un proceso, que registra, la verificación del estado de la cuenta y del saldo disponible, si la cuenta se encuentra en alguna condición distinta a "sin restricción" o si el monto de la compra excede el cupo de la línea de crédito o si el cliente se encuentra fuera de rango, se procede a bloquear la transacción y el cliente debe gestionar en relaciones comerciales (RRCC), la revisión de su situación comercial.

Lo anterior permite determinar que las fuentes de datos son los archivos de los sistemas transaccionales de Venta y los Archivos Maestros de los sistemas de actualización masiva que se ejecutan en procesos nocturnos.

Identificadas las variables claves y las fuentes de datos se procede a gestionar su acumulación en un repositorio con los datos históricos de un periodo representativo, para pasar a la etapa siguiente de limpieza y depuración.

6.3. Limpiar, depurar y completar datos.

Tal como ocurre en muchas empresas, los datos de la operación diaria se encuentran dispersos en distintos sistemas de tratamiento y representación de información y en diversas plataformas de hardware y software, lo cual, naturalmente representa un inconveniente para la recopilación de datos, situación que requiere de procesos especiales para su obtención.

Por otra parte, los datos históricos rescatados, correspondientes a variables no usadas ni disponibles para los sistemas en línea vigentes, habitualmente corresponden a distintos formatos de registros, según sea la versión del sistema transaccional que corresponda al periodo del cual se rescatan los datos.

Lo anterior requiere de un proceso de selección y depuración de datos, para nuestros efectos se recurre a los datos del sistema de control de clientes para obtener los ítems pertenecientes al responsable de la transacción y a los datos del sistema de ventas para los datos de la transacción, y a los datos del sistema de control de existencia para los ítems relacionados con el producto.

Como resultado de esta etapa los datos obtenidos se organizan de la siguiente forma:

ID Atributo	T. Dato	Largo	Detalle
CL_ID	Number	11	Número ique permite relacionarlo con el cliente real.
CL_NCD	Number	5	Numero compras realizadas en un mismo día
CL_LCD	Number	5	Número de locales donde el cliente realizo compras durante el día.
CL_CUPO	Number	8	Monto autorizado para realizar compras
CL_COACTI	Number	11	Código de actividad del cliente
CL_COMUNA	Carácter	11	Código comuna de apertura cuenta
CL_SEXO	Carácter	1	"M" : Masculino; "F" : Femenino
CL_TICREDITO	Carácter	6	L.Crédito: "C" : Cuota; "R" : Refundido
CL_SUCU	Carácter	14	Sucursal apertura cuenta
CL_REGION	Number	8	Región apertura cuenta
CL_EDAD	Number	3	Corresponde a la edad actual del cliente
CL_ANT_CTA	Number	3	Antigüedad cuenta expresada en años
CAT_CL	Number	3	Categoría cliente VIP 1, Normal 0

Tabla 6. 1 Descripción de atributos de la Tabla Cliente.

ID Atributo	T. Dato	Largo	Detalle
CL_ID	Number	11	Número ique permite relacionarlo con el cliente real.
Tx_FechaTx	Fecha	8	Fecha de la transacción
Tx_LC_ID	Number	6	local donde se efectuó la compra.
Tx_POS_ID	Number	11	POS donde se efectuó la venta.
Tx_POS_BOLETA	Number	11	Número de boleta emitido por el POS
Tx_BOLETA	Number	11	Número único de boleta=LC_ID +POS_ID+POS_BOLETA
Tx_tipo	Number	4	Tipo de transacción que se esta efectuando
Tx_TITUL	Number	4	Identificación si tx fue hecha por un titular o un adicional.
Tx_CANT	Number	4	Unidades compradas en la Tx.
Tx_CUOTAS	Number	4	Número de cuotas.
Tx_M_LIQ	Number	4	Monto liquido de compra efectuada en la transacción.
Tx_LPROD	Number	4	Línea de producto al cual corresponde la transacción, esta puede tomar ejemplo los valores : 1)Electrónica, 2)Avances Efectivo 3)Línea Blanca 4)Decohogar 5)Vestuario 6)Zapatería 7)Otros
Tx_NEGOCIO	Number	4	Código de Comercio Asociado a la Tx.
Tx_EN_DISP	Number	1	0 = NO ; = EN DISPUTA

Tabla 6. 2 Descripción de atributos de la Tabla Transacciones.

Los especialistas del área bajo estudio, como parte de su proceso cotidiano, han acumulado datos provenientes de su operación diaria por un período de dos años, manteniendo un archivo histórico de transacciones efectuadas usando como medio de pago las tarjetas de crédito corporativas.

Las características de la base de datos, resultante de este proceso, son las siguientes:

Base de datos con 53.867.644 casos acumulados en el periodo comprendido desde Mayo de 2001 hasta Agosto de 2003, de estos, 176.659 casos son fraudulentos.

La tabla de transacciones de compra tiene registro fijo de largo 162 bytes, con 24 atributos numéricos sin incluir el indicador de

transacción fraudulenta o legal, los atributos que cada registro comprende lo siguiente:

CAMPO	TIPO	LARGO	DESCRIPCION
RUT	NUMBER	10	RUT DEL CLIENTE
DV	CHAR	1	DIGITO
FECHA_TX	DATE	8	FECHA TRANSACCIÓN
MONTO_LIQUIDO	NUMBER	10	MONTO LIQUIDO SE CARGA A LA CUENTA
ABONO	NUMBER	10	MONTO QUE CANCELA AL HACER LA COMPRA
DESCUENTO	NUMBER	10	MONTO DESCONTADO DE LA COMPRA
MONTO_BRUTO	NUMBER	10	MONTO TOTAL DE LA COMPRA
TRANSACCION	NUMBER	2	TIPO DE TRANSACCION (COMPRA,DEVOLUCION,CAMBIO)
LOCAL	NUMBER	4	LOCAL DONDE SE EFECTUA LA TRANSACCION
TERMINAL	NUMBER	4	POS (PUNTO DE VENTA DENTRO DEL LOCAL)
NUMERO_BOLETA	NUMBER	6	NUMERO DE LA BOLETA DE COMPRA
PRECIO_COSTO	NUMBER	10	COSTO PRODUCTO
TITULAR	NUMBER	1	TITULAR/ADICIONAL
DIFERIDO	NUMBER	1	NOMAL/DIFERIDO
FRAUDES	NUMBER	1	MARCA DE FRAUDE
CUOTAS	NUMBER	3	NUMERO DE CUOTAS
HORA	CHAR	5	HORA DE LA TRANSACCIÓN
CODIGO_PRODUCTO	NUMBER	18	IDENTIFICADOR PRODUCTO
CANTIDAD_VENDIDA	NUMBER	2	CANTIDAD DE LOS VENDIDOS
CODIGO_BANDA	CHAR	15	BANDA MAGNETICA
CUENTA	NUMBER	10	CUENTA CLIENTE
CAPTADOR	NUMBER	10	RUT DE CAJERO
SKU	NUMBER	10	STANDARD KEY UNIQUE

Tabla 6. 3 Descripción de los atributos de la tabla de Transacciones.

6.4. Análisis Exploratorio Preliminar

Aplicando técnicas y herramientas provenientes del ámbito de la estadística tradicional, se realiza el análisis exploratorio preliminar, el cual permite determinar que la principal fuente de transacciones fraudulentas provienen de las estaciones expendedoras de combustibles, de farmacias, de lugares de expendio de comida rápida y de recarga de tarjetas de pre-pago telefónico, compras de productos de fácil reducción a dinero en efectivo, estas últimas son tratadas por el analista de riesgo de la empresa bajo estudio, los otros casos de fraudes son derivados a las empresas asociadas responsable de la prestación del servicio.

La comparación de actividad anual, da cuenta de un incremento de la actividad fraudulenta, este análisis nos permite conocer que existe un incremento importante de la actividad fraudulenta el que alcanza al 15%, además existe una mayor presencia de transacciones generadas fuera del ámbito de los puntos de ventas propios de la empresa origen de la muestra, es decir, las que se efectúan en los locales de venta de sus socios comerciales.

A continuación presentamos los resultados del análisis exploratorio de clasificación de datos por líneas de negocios, obtenido de la muestra de datos.

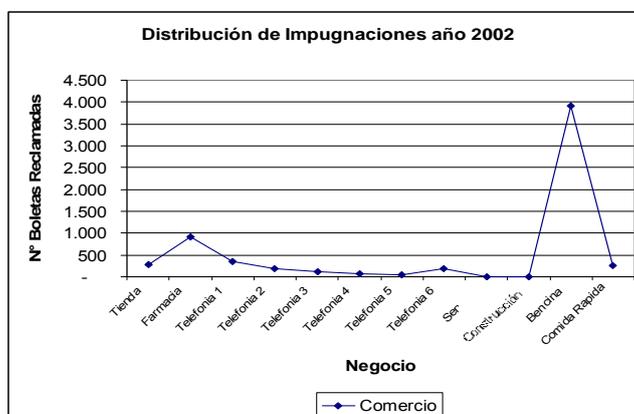


Fig. 6.3 Distribución de Impugnaciones año 2002.

Tal como se observa en la figura anterior, el producto o servicio con mayor cantidad de fraude en el año 2002. Similar situación se presenta en la venta de gasolina, un análisis de la distribución de los domicilios de los clientes sujetos de fraude que viven en la Región Metropolitana señalo lo siguiente:

Las cuatro comunas con mayor número de clientes con actividad fraudulenta se concentran en La Florida con un 11,21%, seguido por Maipú con 8,44%, Las Condes 7,80 % y Puente Alto 7,16. Por el contrario las cuatro comunas con un menor número son Colina con un 0,09 %, seguido por Lampa con 0,13%, Padre Hurtado 0,26% y Lo Espejo con 0,55%.

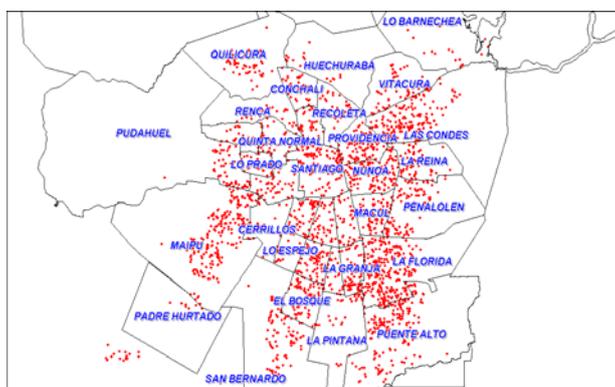


Fig. 6.4 Distribución geográfica de las Impugnaciones

Analizada la misma muestra anterior, con respecto a la distribución por estaciones expendedoras de combustibles, el resultado fue el siguiente:

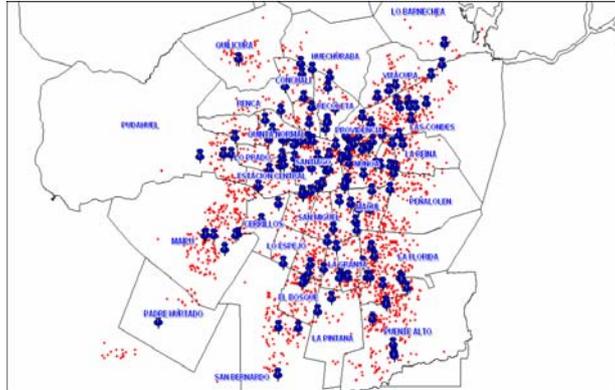


Fig. 6. 5 Distribución de las impugnaciones con respecto a expendedoras de combustible

Revisada la distribución de puntos de ventas propios de esta empresa retail emisora de la tarjeta en el sector donde se produce la actividad fraudulenta, el resultado fue el siguiente:

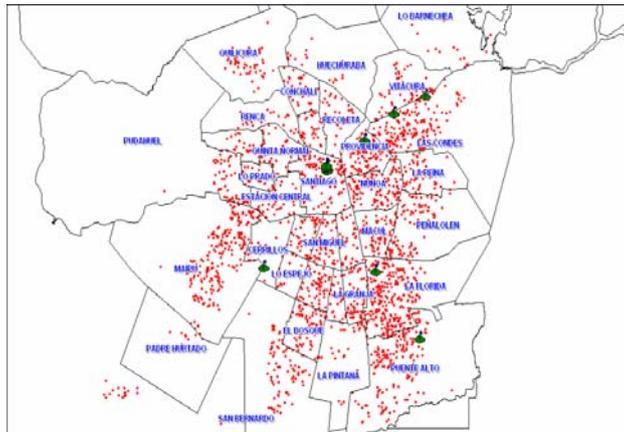


Fig. 6. 6 Distribución de las impugnaciones con respecto a los puntos de venta propios

Analizados los datos de la muestra para determinar la relación entre la ubicación de las tiendas y los centros de expendio de gasolina donde se producen los fraudes, permitió concluir que la relación existe, tal como lo indica la siguiente figura:

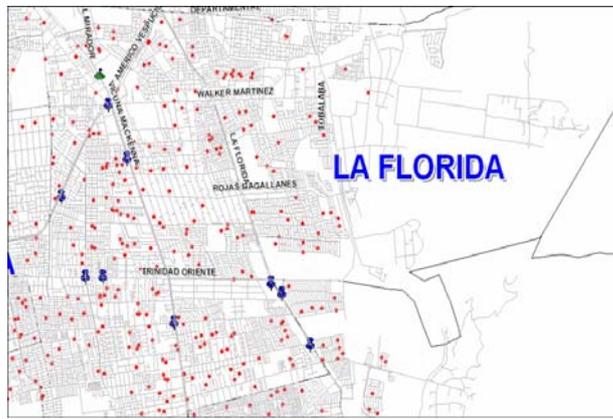


Fig. 6.7 Distribución con respecto a tiendas y expendedoras de combustible

Otro aporte de este análisis estadístico fue la distribución por edad de los tenedores de tarjetas objetos de fraude:

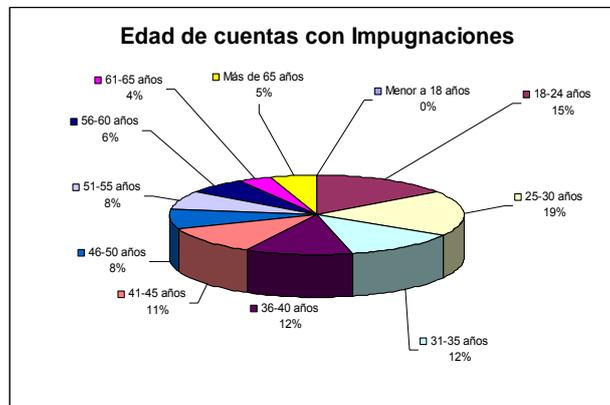


Fig. 6.8 Distribución de las impugnaciones con respecto a la edad

Este gráfico indica la edad de los clientes, 18-30 se concentra un 34%, que comparativamente con las otras agrupaciones de edad se produce una alta incidencia de actividad fraudulenta.

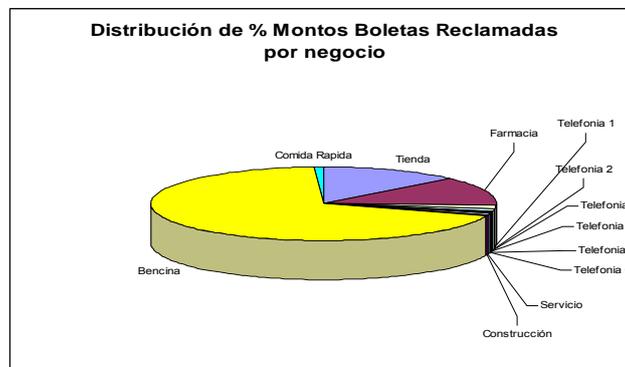


Fig. 6.9 Distribución de Montos

6.5. Interpretación de los resultados

Este análisis estadístico de los datos histórico permitió determinar que la principal fuente de transacciones fraudulentas provienen de las estaciones expendedoras de combustibles (61,6%), de farmacias (14,3%), de lugares de expendio de comida rápida (4%) y de recarga de tarjetas de pre-pago telefónico (15,3), compras de productos de fácil reducción a dinero en efectivo.

En este análisis exploratorio del total de transacciones fraudulentas un 55% corresponden a tarjetas de clientes de sexo femenino, del total de clientes afectados un 34% corresponde al rango de edad entre 18 y 30 años, un 61,6% de estas transacciones fraudulentas corresponde a recarga de combustible, un 15,3 a recargas de tarjetas de prepago telefónico y un 14,3% a compras en farmacias,

Estas transacciones han tenido un aumento con una tasa de crecimiento del 15% anual.

Del análisis estadístico, de las tendencias observadas en la muestra y su discusión con los analistas de riesgo, se desprenden las siguientes reglas o paradigmas que ellos consideran pertinentes:

- El desconocimiento de compra correspondiente a una persona se sexo femenino tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a una persona joven tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a una persona joven tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a un producto de fácil reducción a dinero o servicios de comida rápida, carga de combustible o recarga de tarjeta telefónica tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a una tarjeta de apertura reciente tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a un cambio de conducta del cliente en cuanto a la periodicidad de compra, el lugar donde realiza la compra, o el monto de compra tiene alta probabilidad de corresponder a una situación de fraude.
- El desconocimiento de compra correspondiente a una compra sin pago inicial tiene alta probabilidad de corresponder a una situación de fraude.

Obtenidas las actuales reglas de negocio que aplican los especialistas, se procede a paso siguiente de verificación de la vigencia de las reglas actuales, mediante el análisis comprobatorio.

6.6. Análisis comprobatorio (Parte 1: Extraer las Reglas)

Siguiendo la metodología propuesta, corresponde realizar un Análisis a los métodos, mejores prácticas y herramientas de última generación aplicables al caso bajo estudio.

En este caso, dadas las características de las reglas o paradigmas que considera el proceso de toma de decisiones para determinar si un desconocimiento de compra es o no una situación de fraude y considerando que se están ante un buen ejemplo de dominios inciertos, en donde se puede dar lugar a fuentes de incertidumbre, al uso del lenguaje natural donde la información es esencialmente implícita; donde esta presente la multiplicidad, la ambigüedad y la imprecisión del lenguaje natural, por todo ello, es fundamental incorporar un proceso que trate la incertidumbre, lo cual determina la aplicación de las reglas de asociación difusa, para verificar la vigencia del conocimiento del negocio aplicado a este tipo de decisiones de carácter estratégico para la empresa. El carácter estratégico esta dado por el impacto que causa un desconocimiento de compra en el proceso de fidelización de clientes y la cobertura mediática que situaciones puntuales han tenido en los medios de comunicación masivos.

La aplicación de la técnica de extracción de reglas de asociación difusa en este análisis comprobatorio, requiere desarrollar los siguientes dos procesos:

1) Compatibilizar los datos con las variables y contratos usados por los especialistas, elaborando o definiendo el conjunto de etiquetas lingüísticas que permiten hacer imprecisas las variables difusas de las transacciones de prueba, para incorporarlas a las tablas de clientes y transacciones que constituyen la base de datos bajo estudio.

2) Desarrollar el proceso de extracción de reglas de asociación difusas, procesando las tablas clientes y transacciones con las variables difusas incorporadas en el proceso anterior, para ello se utilizan aplicación de software como FuzzyQuery 2+, de Serrano et al [SJ03] o el software CBA-Aias de Lui [LB99].

6.6.1. Definición de las Etiquetas Lingüísticas

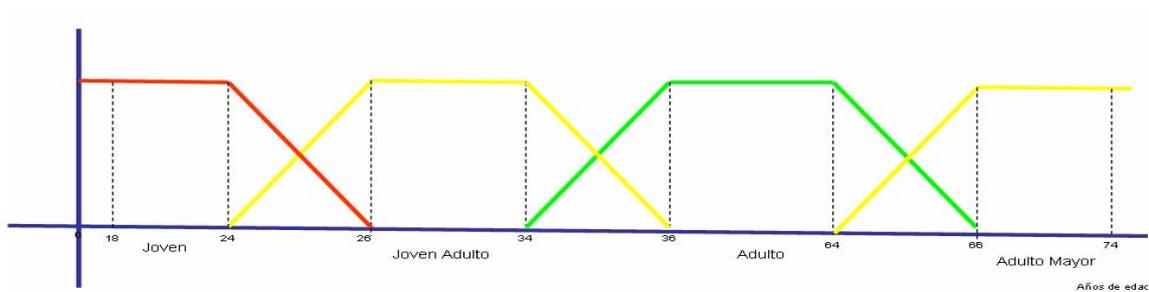
Los procesos anteriores de la metodología propuesta, permitieron determinar las variables claves y los atributos que tienen mayor

incidencia en la determinación de una situación de fraude ante un desconocimiento de compra, estas variables son:

- Tipo de Producto
- Lugar de Compra
- Edad del comprador
- Sexo del comprador
- Antigüedad de la cuenta
- Compra sin abono en efectivo
- Compra en máximo de cuotas
- Plazo de Compra.

Para las variables que presentan vaguedad o imprecisión se generan las siguientes etiquetas lingüísticas:

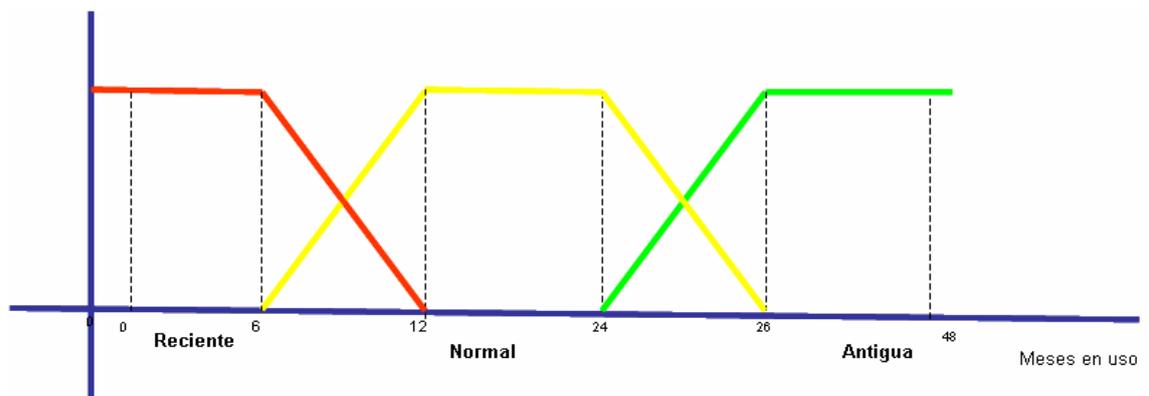
Edad del Cliente genera la siguiente segmentación trapezoidal:



„ Fig. 6. 10 Etiquetas lingüísticas de Edad del Cliente

- Joven (18,18,24,26)
- Joven Adulto (24,26,34,36)
- Adulto (34,36,64,66)
- Adulto Mayor (64, 66,75,75)

Antigüedad de la cuenta (en meses) genera la segmentación trapezoidal siguiente:



„Fig. 6. 11 Etiquetas lingüísticas de Antigüedad de la Cuenta

- Reciente (0,0,6,12)
- Normal (6,12,24,36)
- Antigua (32,40,48,48)

Plazo de Compra en Número de cuotas genera la segmentación trapezoidal siguiente:

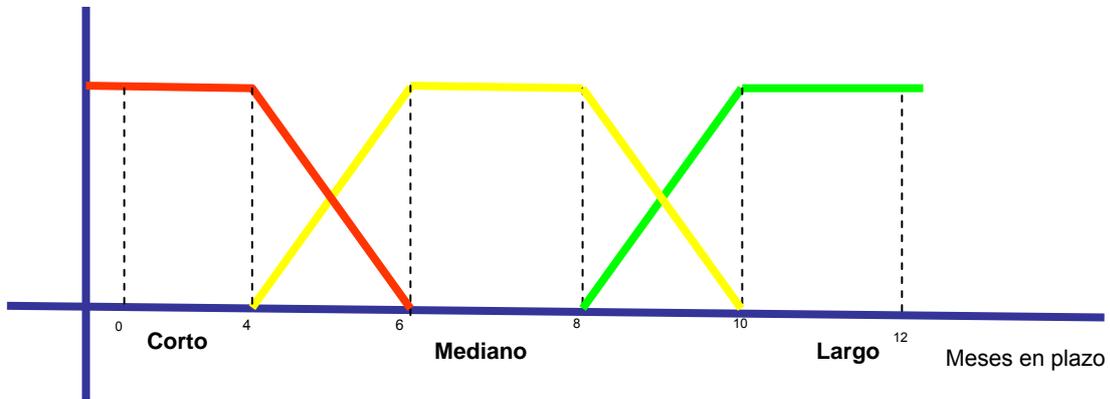


Fig. 6. 12 Etiquetas Lingüísticas de Plazo de Compra

- Corto (0,0,4,6)
- Mediano (4,6,8,10)
- Largo (8,10,12,12)

Monto de la compra (en miles de pesos) genera la segmentación trapezoidal siguiente:

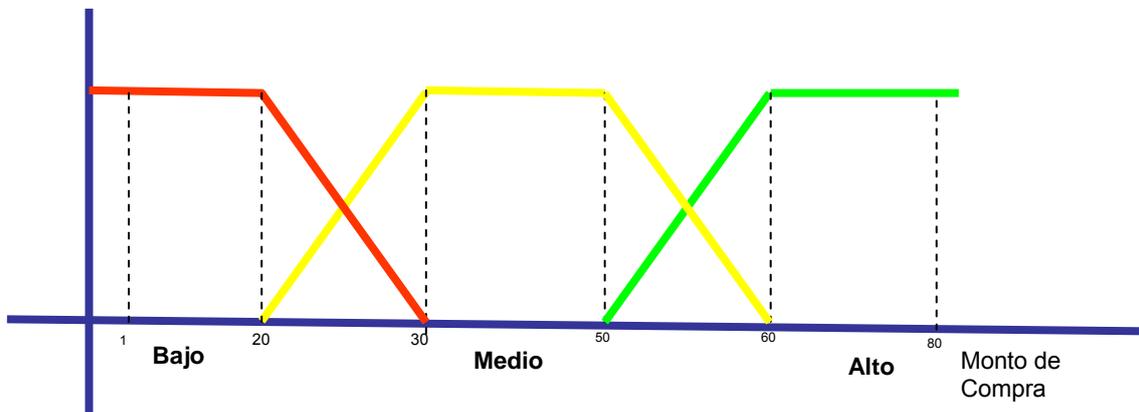


Fig. 6. 13 Etiquetas lingüísticas de Monto de Compra

- Bajo (1,1,20,30)
- Medio (20,30,50,60)
- Alto (50,60,80,80)

6.6.2. Extracción de las Reglas de Asociación Difusa

Una vez definidas e incorporadas las etiquetas lingüísticas a la base de datos, se realiza el proceso de extracción de reglas de asociación difusa, obteniendo los siguientes resultados:

Tabla Clientes:

El experimento se realiza sobre la tabla cliente con 1.959 filas, se aplican los siguientes parámetros:

Nivel de soporte 0.015
Nivel de confianza 60%
Factor de certeza 60%
Número máximo de atributos en el antecedente: 5

Como resultado se obtiene un total de 8.008 reglas de asociación difusa.

Los indicadores del proceso señalan lo siguiente:

436 reglas con dos elementos
1.009 reglas con tres elementos
1.184 con cuatro elementos y
730 con cinco elementos.

Tabla Clientes con la tabla transacciones:

Al experimento realizado con la tabla clientes con 1.959 cuentas se agrega la tabla transacciones con 12.107 filas:

Nivel de soporte 0.015
Nivel de confianza 65%
Factor de certeza 65%
Número máximo de atributos en el antecedente: 4.

Este experimento señala que existe:

600 reglas con 2 elementos
2.429 reglas con 3 elementos y,
5.715 con 4 elementos.

6.7. Análisis comprobatorio (Parte 2: Depurar las Reglas)

Para esta segunda parte del análisis comprobatorio, del conjunto de reglas de asociación difusa extraídas de la tabla clientes, se seleccionan 379 que tienen en el consecuente la variable fraude positivo (estado fraude = 1).

En una revisión preliminar se observa que en estas 379 reglas de asociación difusa, más de 50 tienen un factor de certeza por sobre el 73%.

De igual forma, del proceso conjunto de la tabla cliente y la tabla de transacciones, se genera 19.116 reglas, se observa que más de 50 ellas cuyo consecuente tiene la variable fraude positiva, presenta Factor de Certeza sobre el 66%.

La primera impresión para los especialistas, es la rapidez de obtención de estos resultados, los cuales corresponden a una muestra con un universo de 1.959 clientes que realizaron 12.107 transacciones, lo cual sorprende a los analistas de riesgos y les genera buenas expectativas para continuar con la metodología propuesta.

La metodología propone reducir el número de reglas de asociación difusa, aplicando técnicas de Medidas de Interés Objetiva y Subjetiva.

Las primeras, buscan determinar cuales de ellas son reglas fuertes o muy fuertes, y a ese conjunto someterla al Análisis de medidas subjetivas para finalmente someter a consideración de los especialistas sólo aquellas que efectivamente cumplan con la condición de ser novedosas y útiles.

6.7.1. Aplicación de las técnicas de Medidas de Interés a las reglas obtenidas

Medidas de interés objetivas

El primer proceso de depuración de reglas corresponde a la aplicación de reglas de interés objetivas, proceso que permite reducir el número de reglas a analizar basados en los criterios establecidos por Berzal et al [BF02], que permite seleccionar aquellas reglas de asociación difusas fuertes o muy fuertes.

En consecuencia, se buscan en el conjunto de reglas de asociación extraídas, las reglas muy fuertes, es decir, aquellas que cumplan con $SuppA+SuppC > 1$ y $A \rightarrow C$ es fuerte.

Como resultado del proceso de extracción de reglas de asociación difusa, se generan 4.058 reglas que igualan o exceden los niveles de minSupp y minConf establecidos por el usuario.

Del conjunto de 4.058 reglas, sólo 1.552 reglas cumplen con la condición de ser reglas fuertes, parte de este subconjunto se muestran en orden alfabético, en la tabla 6.5 del anexo F.

Luego, se obtiene del conjunto de 4.058 reglas aquellas que cumplen con la condición de $A+C < 1$, en este caso, 2.506 son menores que 1. Una muestra de ellas se incluye en la tabla 6.6 del anexo F.

Luego, de entre las 2.506 reglas anteriores se extraen las reglas muy fuertes, estas deben cumplir la condición:

Sea $\text{supp}(A)+\text{supp}(C) < 1$. Entonces $A \Rightarrow C$ es muy fuerte si $\neg C \Rightarrow \neg A$ es fuerte.

Algunas de las reglas de asociación difusas que cumplen la condición anterior se muestran en la tabla 6.7 del anexo F.

6.7.2. Medidas de Interés Subjetivas

Obtenido el subconjunto de reglas fuertes y muy fuertes, se aplican los criterios propuesto por Liu et al [LB99], para determinar las medidas de interés subjetivas.

Esta técnica consiste en la aplicación de las jerarquías de similitud, las que se establecen de la siguiente manera:

- a) Jerarquía según la similitud de los ítems del antecedente de las reglas obtenidas en el proceso de minería de datos con los ítems del antecedente de las reglas definidas por el usuario
- b) Jerarquía según similitud de los ítems del consecuente de las reglas obtenidas en el proceso de minería de datos con los ítems del consecuente de las reglas definidas por el usuario
- c) Jerarquía según similitud de los ítems de ambos lados de las reglas obtenidas en el proceso de minería de datos con los ítems de ambos lados de las reglas definidas por el usuario.
- d) Jerarquía según similitud de conformidad de las reglas obtenidas en el proceso de minería de datos con las reglas definidas por el usuario.

A modo de ejemplo, para este proceso se utiliza el conocimiento del usuario que según lo indicado en el apartado 5.1.4, que ha modo de ejemplo señala:

- La mayoría de los casos de fraude son mujeres:
Tipo_ transacción=1 => Sexo="Femenino"
- La mayoría de las personas sujetas de fraude son jóvenes
Tipo_ transacción=1 => Edad="Joven"
- Los principales productos involucrados en fraude son de fácil reducción a dinero
Tipo_ transacción=1 => Producto="ReducibleEnDinero"

- Un gran porcentaje de las tarjetas involucradas en situaciones ilícitas son de reciente apertura:
Tipo_ transacción=1 => Tarjeta="Reciente"
- Un gran porcentaje de las situaciones ilícitas son realizadas en el sector sur:
Tipo_ transacción=1 => Tienda="Sector Sur"

a) Resultados del proceso de determinar la jerarquía según la similitud de los antecedentes

Al analizar la presencia de los ítems en el antecedente o atributo en el costado izquierdo de las reglas muy fuertes se determinan situaciones como las siguientes:

La afirmación:

Si es Mujer y Joven entonces Fraude es positivo,

Este paradigma se rompe en la muestra bajo estudio, donde es más relevante la combinación Hombre y Joven cuyo soporte alcanza al 15,7%, según se aprecia en la siguiente tabla 6.4. También es evidente que la mayor presencia se da en compras en tiendas y efectuando compras únicas, situación que es aceptada por los especialistas como un nuevo paradigma dejando obsoleto aquel que señalaba que eran las mujeres jóvenes las más afectadas con múltiples compras.

A1	A2	A3	=>	C1	C2	C3	Supp	Conf	Fcert
Hombre	Joven		=>	NroLocalBajo	Fraude		15,741	100,000	1,000
Hombre	Joven		=>	CompraTienda	Fraude		15,741	100,000	1,000
Mujer	Joven		=>	NroLocalBajo	Fraude		11,111	100,000	1,000
Mujer	Joven		=>	CANT1	Fraude		11,111	100,000	1,000
Mujer	Joven		=>	CompraTienda	Fraude		11,111	100,000	1,000
Mujer	Joven		=>	TX_TITUL = 1	Fraude		10,185	91,667	0,700
Mujer	Joven		=>	CortoPlazo	Fraude		8,333	75,000	0,518
Hombre	Joven		=>	NroLocalBajo	Fraude		7,407	100,000	1,000
Hombre	Joven		=>	CompraTienda	Fraude		7,407	100,000	1,000
Mujer	Joven		=>	NroLocalBajo	Fraude		5,556	100,000	1,000
Mujer	Joven		=>	CANT1	Fraude		5,556	100,000	1,000
Mujer	Joven		=>	CompraTienda	Fraude		5,556	100,000	1,000

Tabla 6.4 Ejemplo de Análisis de jerarquía según similitud de antecedente.

b) Resultados del proceso de determinar la jerarquía según la similitud de los consecuentes

Al analizar la presencia de los ítems en los consecuentes o atributo en el costado derecho de las reglas muy fuertes se determinan situaciones como las siguientes:

La afirmación

Si son fraudes de bajo monto entonces se compra varias veces hasta agotar el cupo disponible,

Este análisis de los consecuentes señala que las compras diarias únicas son mayoritarias según lo indica la tabla 6.5, con la cual, los especialistas deciden reemplazar este paradigma con esta nueva evidencia.

A1	A2	A3	=>	C1	C2	C3	Supp	Conf	Fcert
LargoPlazo	MontoBajo	Fraude	=>	CompraDiaria1			26,852	100,000	1,000
LargoPlazo	MontoAlto	Fraude	=>	CompraDiaria1			6,481	100,000	1,000

Tabla 6.5 Ejemplo de Análisis de jerarquía según similitud de los consecuentes.

c) Resultados del proceso de determinar la jerarquía según la similitud de ambos lados de la regla

Luego al analizar la presencia de los ítems indicados por el usuario, en ambos lados de la proposición, la ocurrencia de ítems en el antecedente y en el consecuente, se obtienen resultados como el siguiente ejemplo que corresponde a la afirmación:

La combinación Mujer + Joven => Fraude, es mayor que combinación Hombre + Mujer => Fraude

En la tabla 6.6, se observa lo contrario, que la combinación Hombre+Joven es superior que la combinación Mujer+Joven, con esta evidencia, los especialistas, también deciden reemplazar este paradigma.

A1	A2	A3	=>	C1	C2	C3	Supp	Conf	Fcert
Hombre	Joven		=>	NroLocalBajo	Fraude		15,741	100,000	1,000
Mujer	Joven		=>	NroLocalBajo	Fraude		11,111	100,000	1,000
Hombre	Joven		=>	CompraTienda	Fraude		15,741	100,000	1,000
Mujer	Joven		=>	CompraTienda	Fraude		11,111	100,000	1,000

Tabla 6.6 Ejemplo de Análisis de jerarquía según similitud de ambos lados.

d) Resultados del proceso de determinar la jerarquía según la conformidad de las reglas

Finalmente la técnica que completa lo propuesto por Liu et al [LB99], verifica el grado de conformidad de las reglas obtenidas en este proceso con las expectativas del usuario, determinando el grado de conformidad cuantificando si el número de elementos que conforman el antecedente y el consecuente son los mismos que conforman la definición entregada por el usuario, estableciendo el grado de similitud entre ambas.

De este proceso, por ejemplo, la regla que tiene mayor puntuación es la que se muestra en la tabla 6.7, que señala que la mayor parte de los fraudes se realiza haciendo una única compra:

A1	A2	A3	=>	C1	C2	C3	Supp	Conf	Fcert
CANT1			=>	CompraTienda	Fraude		97,222	100,000	1,000

Tabla 6.7 Ejemplo de Análisis de jerarquía según conformidad de las reglas.

Lo anterior contradice el paradigma que indica que se realizan varias compras de bajo monto hasta agotar el saldo disponible.

Por otra parte, este proceso, también permite confirmar que los fraudes donde se adquiere productos de fácil reducción a dinero se da por igual para ambos sexos, como se indica en la tabla 6.8, que muestra la presencia similar para ambos géneros.

A1	A2	A3	=>	C1	C2	C3	Supp	Conf	Fcert
Hombre	DeFacilReduccion		=>	CANT1	Fraude		9,259	100,000	1,000
Hombre	DeFacilReduccion		=>	CompraTienda	Fraude		9,259	100,000	1,000
Mujer	DeFacilReduccion		=>	CANT1	Fraude		9,259	100,000	1,000
Mujer	DeFacilReduccion		=>	CompraTienda	Fraude		9,259	100,000	1,000

Tabla 6.8 Ejemplo de Análisis de jerarquía según conformidad de las reglas

6.8. Incorporar los resultados a la lógica del negocio

Para concluir la aplicación de la metodología Fuzzy Business Intelligence que permite verificar la vigencia y actualizar el conocimiento del negocio, se debe incorporar los resultados obtenidos a la lógica del negocio.

En este caso, el proceso se realiza con los expertos, quienes al analizar los resultados de los experimentos realizados, concluyen que de las 21 reglas de asociación difusa seleccionadas, existen 15 adicionales a las de uso habitual.

De esas 21 reglas, varias constituyen una novedad, por ejemplo, las que rompen paradigmas, como que la mayoría de los fraudes afectan a mujeres, el experimento indica, para esta muestra, que la combinación sexo, región y tarjeta reciente que tiene una presencia en la muestra del 2,6%, tiene una confianza y un Factor de Certeza 92%, de igual forma los fraudes en Santiago Centro tienen mayor relevancia que las que ocurren en el Mall Vespucio.

Hecho este proceso y dado que en este caso, la organización presenta un nivel de madurez de informática de gestión incipiente, son los mismos especialistas los que adoptan los nuevos paradigmas, y realizan por ejemplo lo siguiente:

Eliminan de sus reglas de negocio las que indican un cambio de paradigma, como por ejemplo aquella que señala que un desconocimiento de compra de una mujer joven es potencialmente un fraude y aquella que señala que un potencial fraude es aquel que se realiza reiterando compras pequeñas hasta agotar el cupo.

Modifican las reglas de negocios, por ejemplo introduciendo que las transacciones que causan desconocimiento de compra provienen en su mayor parte de Santiago Centro en lugar de La Florida.

Actualizaciones como las señaladas han permitido a los especialistas mantener vigente el conocimiento del negocio en la empresa y les otorga una herramienta eficiente para que actualicen sus paradigmas.

6.9. Resumen del Análisis del Proceso de Gestión de Riesgo.

Este capítulo presentó la aplicación de la metodología Fuzzy Business Intelligence al área de gestión de riesgo y en particular a las decisiones estratégicas relacionadas con los desconocimientos de compras.

El resultado de este proceso ha mostrado efectividad de la metodología propuesta, en cuanto a su potencial entregando a los analistas de riesgos conocimiento útil e inesperado, permitiendo actualizar la lógica de negocios con el conocimiento vigente extraído desde las bases de datos históricos de la empresa bajo estudio.

Capítulo 7

Proceso de Gestión de Negocios

Caso: Proceso de Gestión Estratégica del Negocio.

Este capítulo presenta la segunda aplicación de la metodología Fuzzy Business Intelligence al proceso de gestión de negocios, cubre uno de los procesos de decisiones del área de Gestión de Negocios, encargada de realizar los Procesos de Planificación, Presupuesto, Compra y Control de Resultados de la Venta al Detalle.

Las motivaciones para aplicar la metodología Fuzzy Business Intelligence al área de Gestión de Negocios, entre otras, son las siguientes:

En el área de Gestión de Negocios de la empresa bajo estudio, las decisiones estratégicas se toman en términos generales bajo condiciones de incertidumbre, derivadas de información errónea, incompleta e imprecisa, en una realidad no determinista y aplicando modelos incompletos o inexactos, esto significa que ésta área, presenta las típicas condiciones de los ejemplos paradigmáticos de dominios inciertos para aplicar la metodología Fuzzy Business Intelligence.

En el área bajo estudio, el uso de argumentos en lenguaje natural es habitual para justificar una decisión estratégica, esto significa que la información implícita, la multiplicidad, la ambigüedad y la imprecisión contribuyen a la incertidumbre en la toma de decisiones.

El volumen de datos históricos acumulados y la carencia de recursos tecnológicos para su almacenamiento y tratamiento, como también la presencia de procedimientos semi-manuales, basadas en herramientas de análisis de datos que usan las funcionalidades de planillas electrónicas y cubos de datos con alto nivel de granularidad, permiten mantener vigente el conocimiento sólo sobre el 20% de los productos.

7.1. Determinar el perfil de la Empresa y el Modelo de Negocios

El perfil de la empresa del segundo caso bajo estudio corresponde a la empresa líder en el segmento de mercado que esta posicionada, el análisis del macro entorno y de la industria, al igual que el caso anterior, se detalla en los apartados 4.1 al 4.6 y de manera más detallada en el anexo D.

El modelo de negocios corresponde al de distribución minorista que al ubicarse entre la producción y el consumo, crea *utilidades* a los consumidores y *servicios* a los productores. La distribución crea al consumidor utilidad de lugar, de tiempo, de forma y de creación de surtidos y de posesión estas utilidades creadas en el mercado de vestuario y hogar para los consumidores se relacionan en forma directa con los servicios que crea para los productores (transporte, almacenamiento, "finalización del producto", información, financiación y asunción de riesgos).

En el modelo de gestión de negocios que utilizan las empresas subyacen los modelos que utilizan los indicadores de gestión relacionados con la rentabilidad de las inversiones en inventarios, con la rentabilidad de los metros cuadrados de exhibición, y la rentabilidad de los gastos de administración y venta. Todo esto con la habida consideración de la adecuada gestión del surtido y precio del producto.

7.2. Identificar las variables claves y las fuentes de datos.

Para determinar las variables claves, la metodología recomienda revisar los procesos que realizan los especialistas, en la siguiente figura 7.1, se muestra el actual modelo de toma de decisiones en el proceso de negocios de una de las grandes tiendas bajo estudio:

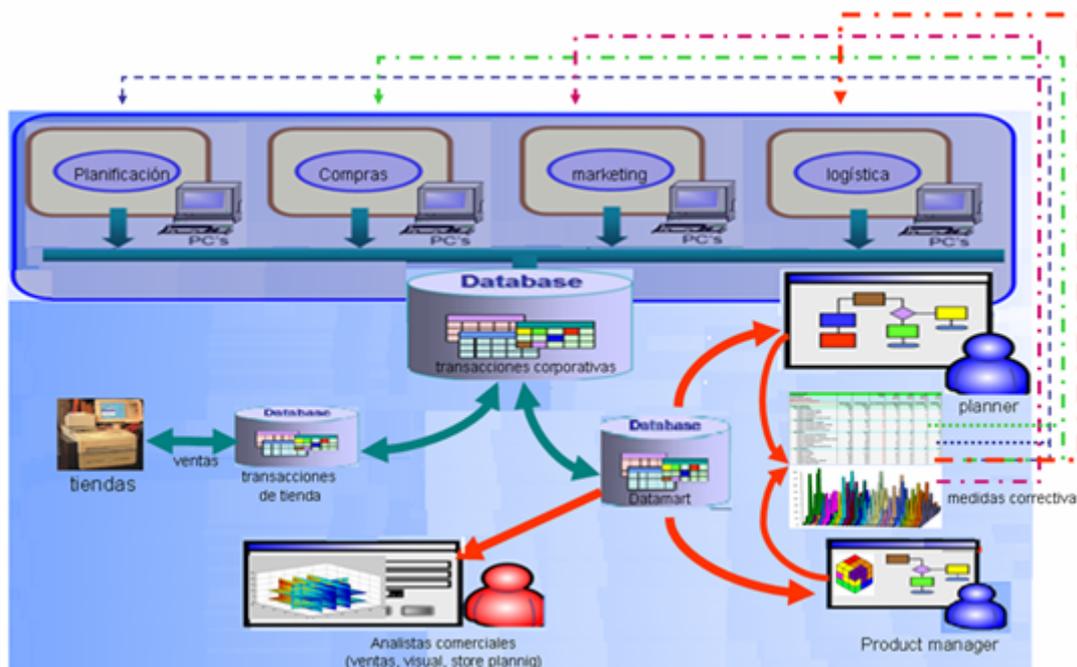


Fig. 7.1 Actual Modelo de toma de decisiones

En la figura 7.1, se verifica que la administración de las medidas correctivas (parámetros) para la estrategia de negocios esta basada en el uso de métodos estadísticos tradicionales y en la intuición de los analistas (planners). Las medidas correctivas que los planificadores y los gerentes de productos toman son el resultado de análisis del

comportamiento de las variables que intervienen en el modelo de negocios y que las obtienen de la información que provee el datamart con serios problemas de granularidad ya que solo permite conocer datos hasta el nivel de artículos. Obtener datos a nivel de ítem requiere realizar procesos especiales nocturnos sobre las bases de datos transaccionales. Obtenidos los datos, se traspasan a planillas Excel, donde se aplican técnicas estadísticas básicas que les permiten evaluar y tomar las decisiones las que son implementadas a través de la modificación de parámetros y datos del sistema informático transaccional, medidas que computacionalmente tienen efecto al día siguiente de ser ingresadas.

En la actualidad los analistas de negocios destinan más del 60% del tiempo en la recopilación de los datos y menos del 40% en las labores de análisis, lo que les permite concentrar su atención en un máximo del 20% de los productos, recibiendo atención aquellos que generan el 80% del ingreso de la línea de negocios a su cargo.

Los indicadores más usados en este ámbito son el GMROI y como referencia el GMROS y el GMROL.

Los Modelos de referencia que utilizan los analistas de negocios son el de Gerenciamiento de Ingresos y Precios y el Gerenciamiento de las Categorías ambos descrito en el Capítulo 2.

En este sentido, una herramienta eficaz para consolidar ambos modelos es la cantidad óptima a exhibir (que corresponde a una de las acepciones del concepto Kanban).

El Kanban debe ser aumentado cuando:

- Existen Eventos especiales
- Existe Alta rotación
- Ante un aumento de ventas
- Alta disponibilidad
- proximidad de un período de alta demanda.

Por otra parte el Kanban disminuye cuando:

- Producto quebrado sin existencia en bodega para reponer esto implica consolidar en tiendas de mejor venta
- Ciclo de vida cumplido
- Termino de evento comercial
- Artículo no tiene buena venta, es decir, se encuentra "Pegado".
- Artículo perdedor
- Bajo nivel de venta
- Termino de temporada
- Devolución a proveedor

Variables consideradas importantes para administrar el óptimo de exhibición (Kanban) son las siguientes:

- Comportamiento y Tendencias Climáticas (temporada normal o fuera de lo normal (muy lluviosa; muy seca; mucho calor, fría, etc).
- Eventos programados (Eventos programados como liquidaciones; venta nocturna; Campaña de Avisaje Publicitario, etc).
- El artículo forma parte o complementa una promoción (2x1; el segundo gratis; por la compra lleve de regalo...)
- Cantidad Disponible en bodega para distribuir a tienda
- Nivel de la venta promedio (nivel de Rotación)
- Tipo de tienda (ganadora, perdedora), situación particular de cada tienda
- Estacionalidad
- Historia de venta (estimaciones)
- Etapa del ciclo de vida del producto

Algunos ejemplos de Reglas de Negocios actualmente en uso:

1. Si "la carga es baja" y "la venta es buena" o "el producto es ganador" y "la tienda es ganadora" y "la tienda tiene espacio disponible" y "en bodega hay disponible" entonces "aumentar la cantidad en exhibición (kanban)"
2. Si "esta programado un evento promocional" y "no hay carga suficiente" disponible" y "existe espacio disponible en tienda" entonces "aumentar la cantidad en exhibición (kanban)"

Por otra parte algunas de las reglas de negocio reactivas son:

1. Si "el producto esta quebrado" y "la venta no es buena" entonces redistribuir o concentrar productos donde "el producto es ganador" y "la tienda es ganadora" y "la tienda tiene espacio disponible", esto significa: "disminuir el kanban en la tienda perdedora" y "aumentar la cantidad en exhibición (kanban) en la tienda ganadora"
2. Si "es fin de temporada" o "ha terminado la etapa plena del ciclo de vida del producto" y "la venta no se dio" entonces "enviar el producto a liquidadora" y "dejar el kanban en cero".

En consecuencia, entre otras, las variables que eventualmente se tendrían en cuenta para extraer reglas de asociación son las siguientes:

1. Área de negocios
2. Nivel de existencia en la tienda (Stock)
3. Ciclo de Vida artículo

4. Planificación o Vigencia de Eventos Promocionales
5. Condiciones o tendencias Climáticas de la estación
6. Clasificación del Artículo
7. Existencia de productos según la combinación talla-color (Curva Completa del artículo en tienda)
8. Comportamiento de la venta (Días Venta)
9. Existencia de productos en bodega central (Disponible para distribuir)
10. Fecha de la última modificación del óptimo de exhibición (Kanban)
11. Comportamiento temporada anterior (Historial de Venta)
12. Nivel óptimo de exhibición (Kanban)
13. Margen de Contribución
14. Nivel de Rotación (Días Venta)
15. Capacidad de exhibición en tienda para la prenda (Perfil)
16. Tipo de Prenda, de acuerdo a ventana
17. Prioridades de Reposición por tienda (algoritmo despacho desde bodega)
18. Ranking del producto (producto ganador, perdedor, promedio)
19. Ranking tienda (tienda ganadora, perdedora, promedio)
20. Grado de avance de la Temporada (inicio, pleno, liquidación)
21. Tipo o Característica de la Tienda (Pequeña, mediana, grande)
22. Tipo Evento
23. Tipo Venta
24. Geo-ubicación de la tienda
25. Velocidad reposición bodega (según ubicación geográfica)
26. Velocidad reposición proveedor (proveedor local)
27. Venta esperada en unidades por temporada (Según Presupuesto)

Dadas todas estas condiciones, una de las labores relevantes en las decisiones estratégicas es:

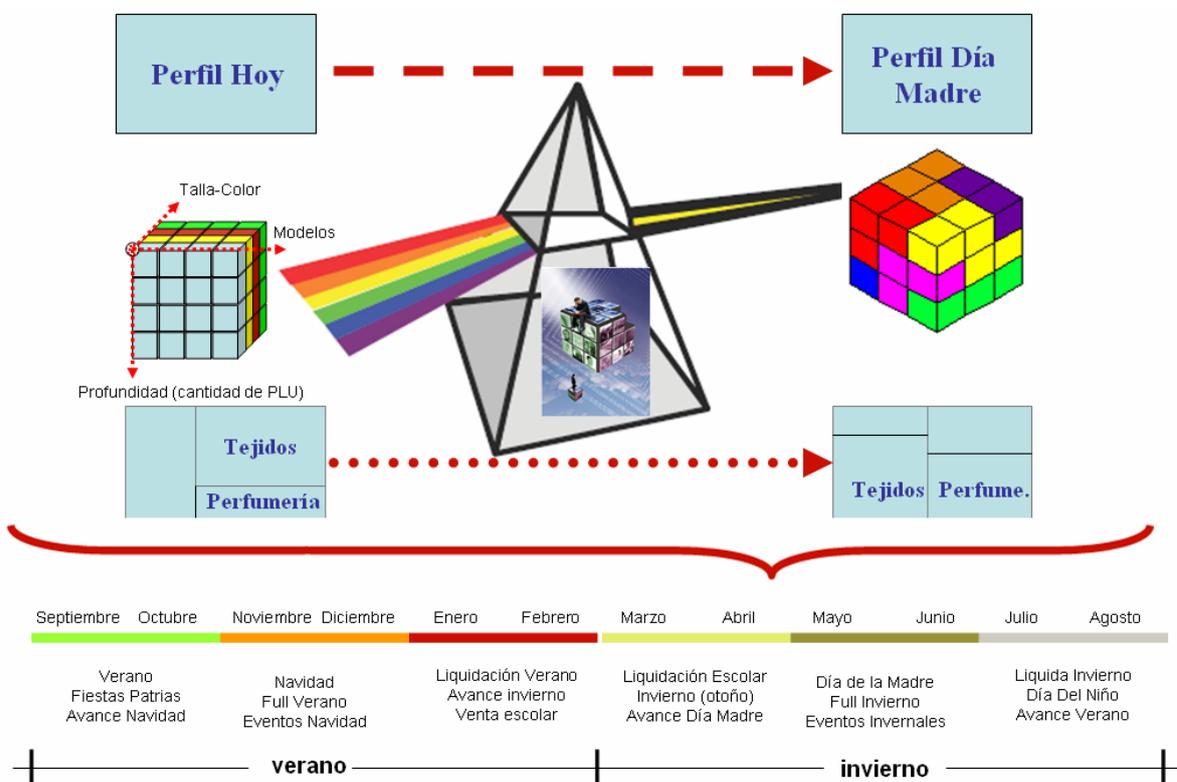
1. A partir de la información de mercado, determinar las nuevas tendencias de la moda.
2. Determinar la cantidad de mercadería a comprar a partir del informe de tendencias y la información histórica de la demanda en cada temporada.
3. Determinar la estrategia competitiva macro con el calendario de eventos y promociones anuales con las cuales se realizaran los llamados a los clientes para anunciarles las ofertas que la cadena pone a su disposición.
4. Establecer las políticas y procedimientos promoción y de ingreso y retiro de mercadería de la tienda, para producir una renovación permanente de la oferta exhibida en la tienda. Reponiendo los productos con buena demanda, empujando con promociones la venta de productos sin

demanda y retirar aquellos que a pesar de las acciones anteriores no fueron demandados, liberando este espacio para el ingreso de nuevos productos.

Para la renovación de la exhibición, se define una capacidad máxima para exhibir una determinada prenda (perfil), esta capacidad se divide de acuerdo a la demanda estimada entre todos los PLU que conformarán el perfil. De esta forma a cada PLU se asocia una cantidad máxima de exhibición (kanban), y la sumatoria de estos kanban constituye el Perfil de la prenda.

De esta forma el kanban se constituye en el elemento regulador de la cantidad de PLU que serán exhibidos, pudiendo aumentar para tener más presencia en la sala de venta o disminuir liberando ese espacio para el ingreso o aumento de exhibición de otro PLU con mejor venta o uno nuevo que se integra a la exhibición.

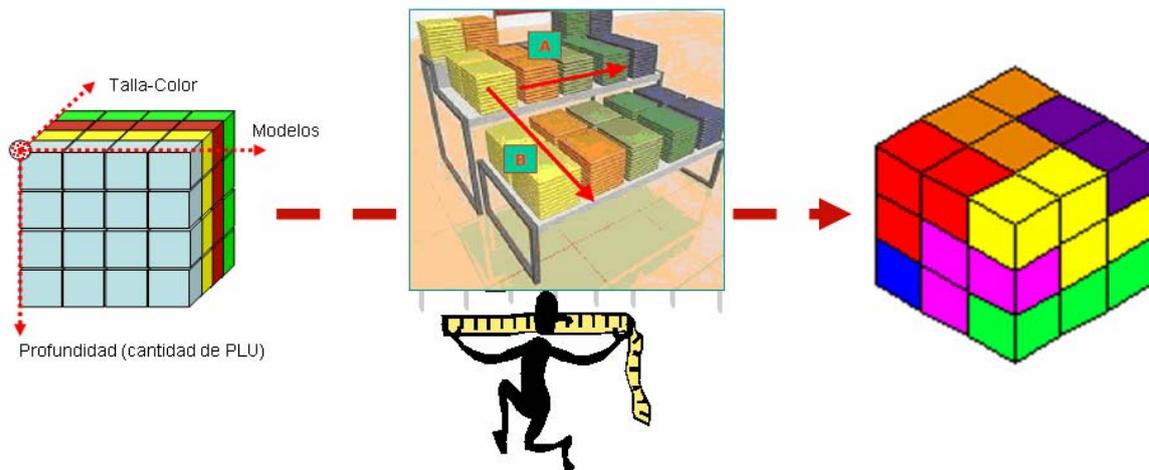
La figura 7.2, siguiente muestra un ejemplo del proceso de ajuste del perfil en el tiempo, donde existen dos temporadas (invierno y verano), y un conjunto de etapas (avance de temporada, temporada, plena temporada, liquidación), y un conjunto de eventos promocionales, todo lo cual, debe ser compatible con la exhibición de artículos, en el ejemplo, corresponde a la festividad del Día de las Madres, en la cual, los productos de tejidos y de perfumería tienen una participación relevante:



.. Fig. 7. 2 Modificación del Perfil.

Para lograr el efecto deseado se utiliza la modificación del Kanban por PLU de cada tienda, aumentando o reduciendo esta cantidad, de manera de provocar un mayor o menor ingreso del PLU desde la bodega a tienda, garantizando su presencia o retiro de la exhibición, logrando transformar la tienda para que presente características coherentes con el apoyo visual en tienda, publicidad en medios de difusión masiva (revista institucional, prensa, radio y televisión).

Para esta adecuada gestión de la capacidad de exhibición por prendas o secciones de la tienda (perfil), el proceso se inicia con el presupuesto de temporada, luego la planificación de compras, recepción, despacho, ingreso, venta y retiro de mercadería de la tienda, proceso en el cual se tienen en cuenta, tal como lo grafica la siguiente figura 7.3, los factores de talla-color, modelos y profundidad de cada PLU logrando así una adecuada gestión de categorías en cada área de exhibición, en este proceso, el concepto del perfil y el kanban juegan un rol preponderante.



.. Fig. 7.3 Gestión de la Capacidad de Exhibición.

Este proceso se realiza con todas las áreas de todas las líneas de negocios, y tal como lo indica la siguiente figura 7.4, da lugar a una tienda con una exhibición agradable, coherente y conveniente para el cliente.

Layout de Tienda – Áreas - Muebles

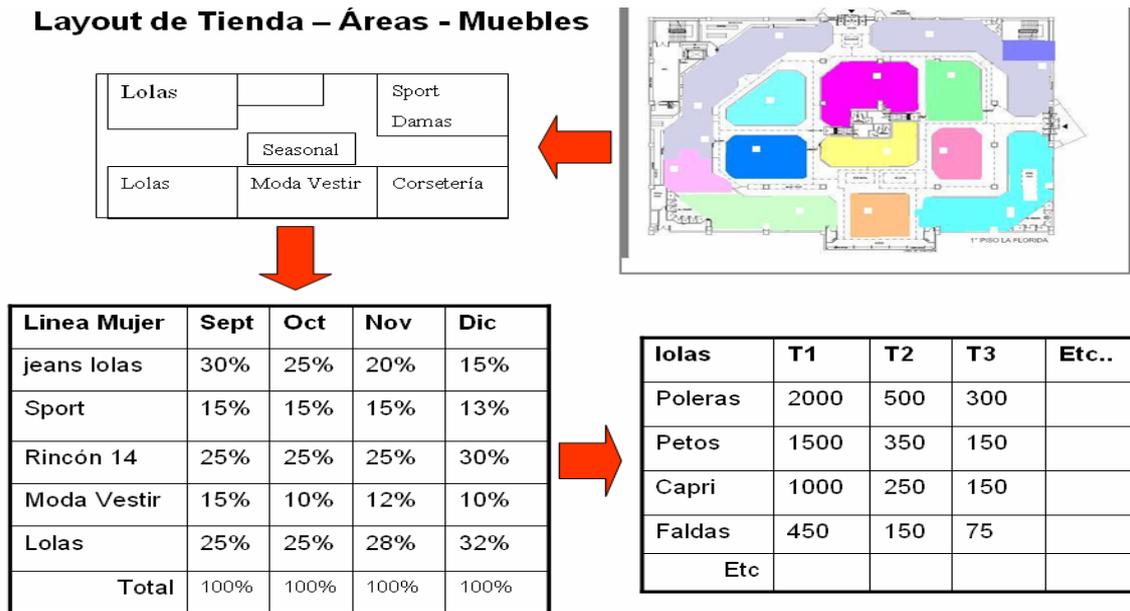


Fig. 7. 4 Adaptación del Layout de tienda.

Como es habitual en cualquier empresa, los datos de la operación diaria y los necesarios para la gestión se encuentran dispersos en distintos sistemas de tratamiento y representación de información, en el ámbito bajo estudio algunos de ellos residen en las bases de datos corporativas, bases de datos de gestión y bases de datos de los puntos de ventas, otros deben ser deducidos o calculados a partir de los datos históricos existentes, lo cual genera subprocesos que permitan corregir las imperfecciones de los datos. En el caso bajo estudio, se recurre a la información histórica de Productos, Inventario, Locales o Tiendas, y al de movimiento de existencias y el histórico de ventas y de datos de gestión.

Una vez identificadas las fuentes de datos se da comienzo al proceso de extraer, limpiar y depurar los datos, estos procesos quedan graficados en la figura 7.5, siguiente y las tablas utilizadas son las siguientes:

	Tabla	Registros
1	Inventario	1.437.912
2	Productos	296.022
3	Histórico de Ventas	40.937.937

Tabla 7. 1 Tablas utilizadas en el experimento.

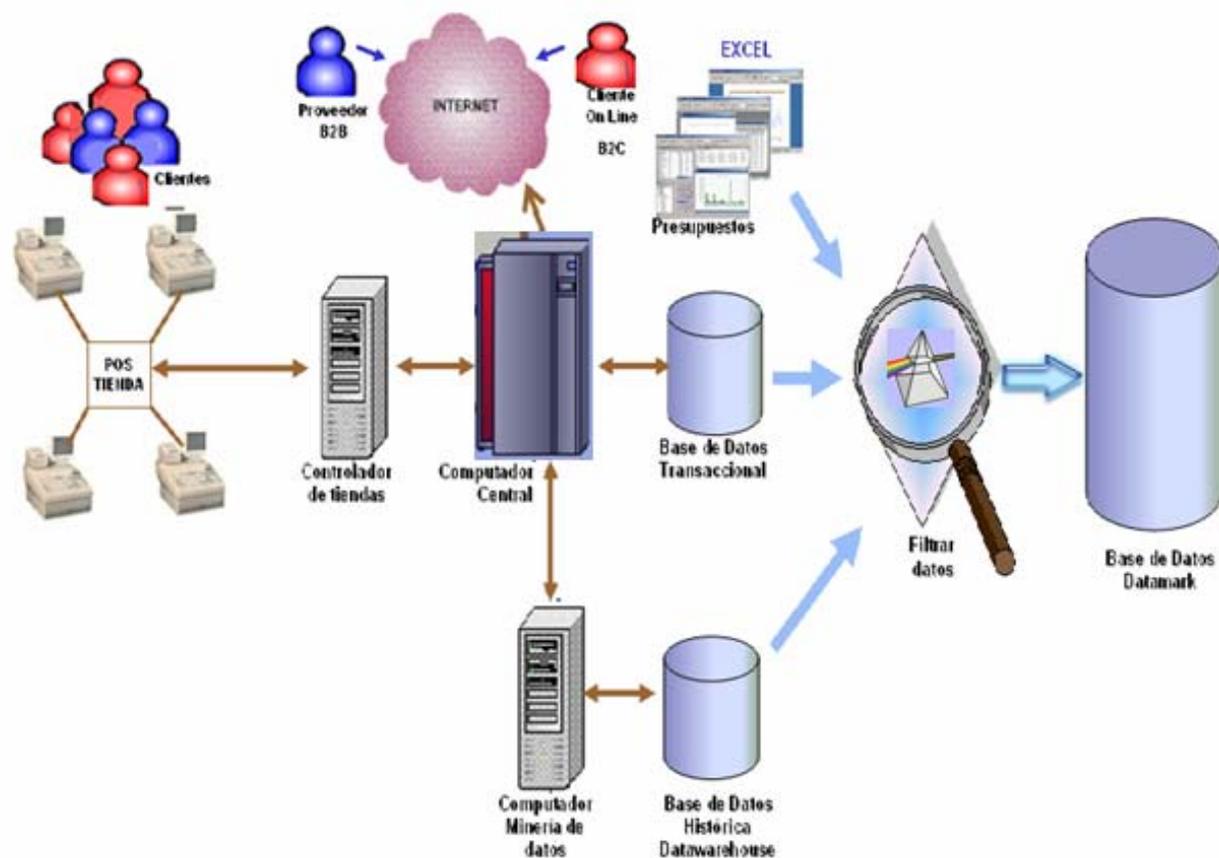


Fig. 7. 5 Fuentes de datos.

Desde las cuales se obtienen los siguientes datos que constituyen la muestra de datos de trabajo:

Variable	Descripción
IdLocal	Corresponde a la identificación de la tienda que efectuó la venta
CodLinea	Identificador de la Línea de Negocios, por ejemplo: Mujer, Hombre, Infantil, Deporte, etc.
CodArea	Identificador del área dentro de la línea, por ejemplo en Mujer: Lencería, Sport Casual; Moda Vestir; Corsetería; Accesorios; Adolescente; Perfumería, etc.
CodSección	Identificador de la Sección, por ejemplo dentro de Sport Casual se tienen: Blusa; Falda; Pantalón; Polera; Tejidos; Vestidos, etc.
CodRubro	Identificación de Rubro o Prendas, en el caso de las poleras se tiene: Vestir; Sport; Tejida
CodArticulo	Identifica al Artículo dentro del Rubro, por ejemplo: Polera Manga Larga Fashion
Plu	PLU o SKU: Numero asignado dentro de la línea a cada combinación talla-color del artículo, corresponde al último nivel de agregación de la mercadería
Color	Color del Artículo
Talla	Talla del Artículo
Origen	Procedencia: Nacional o Importado
Fecha_Activación	Fecha en la cual el producto debe comenzar a exhibirse en tienda.
Fecha_Retiro	Fecha en la cual el producto será retirado de tienda, expirando su ciclo de vida.
FechaVta	Fecha de evaluación: corresponde a la fecha de la transacción venta.
VentaDia	Unidades vendidas en la fecha y local de venta
StockIni	Corresponde al Stock Inicial en la tienda al comenzar el día, es equivalente al stock con que cerró la tienda el día anterior
DiaVenta	Corresponde al cociente entre el StockIni y la VentaDia (índice de rotación de inventario considerando un día de venta)
DiaVentaProm	Corresponde al promedio ponderado de los DiaVenta de los últimos 30 días en que no hubo quiebre de stock para este PLU

Reposición	Cantidad de reposición del PLU para la fecha de la venta
Kanban	Cantidad a exhibir en esta tienda del PLU
EstadoPLU	Codificación según contribución del PLU dentro de los productos del área: A = artículo dentro de los que aportan el 80% de la Contribución del área (venta-costo) B = PLU similar al criterior anterior y que corresponde al siguiente segmento de contribución, PLU se encuentra entre los que aportan el 80% del 20% restante del área C = Resto del los PLU que no forman parte de la clasificación A o B. D = Descontinuado R = Revista (Importado 3 semanas; Nacional 1 semana) P = Prioridad (por algún criterio)

„Tabla 7. 2 Descripción de las Variables Seleccionadas

7.3. Limpiar, depurar y completar los datos.

Determinadas las variables, los constructor y las fuentes de datos, se procede a recopilar los datos desde sus distintas fuentes, dado el volumen que implica, se determina de común acuerdo con los analistas, recopilar datos de una prenda que aporta ingresos significativos a la línea de negocios (prenda ganadora) de una línea de negocios que genera presencia de clientes (línea mujer) en un período de tiempo acotado para tener posibilidad de comparar los resultados una vez modificado el modelo de gestión (temporada verano 2005-2006).

Los datos generales de la muestra de datos son los siguientes:

Línea de Negocios : Línea Mujer
Área : Sport Casual
Sección : Polera Sport
Artículo : Polera Importada
Tipo : Masivo
Origen : Oriente
Temporada : Verano
Variedad Talla-Color : 16 variedades (PLU o SKU)
Dependencias : 10 tiendas (30% del dominio)
U. Comercializadas : 6.400.

Los datos específicos de la muestra de datos son los siguientes:

Origen de los Datos:

Los datos preliminares considerados para realizar el primer análisis exploratorio y revisar el comportamiento de la prenda bajo estudio fueron obtenidos de las siguientes tablas de las bases de datos operacionales:

Ámbito : Control de Transacciones del Periodo
Periodo : desde el 01/08/05 y hasta el 04/09/06

Tabla Histórica de Transacciones del período

Dependencias : Todas
Granularidad : PLU
Registros : 40.937.937 de Transacciones.

Tabla de Control de Inventarios de Tiendas (Stock)

Dependencias : Todas
Granularidad : PLU
Registros : 1.437.912 Filas

Tabla Maestra de Productos

Granularidad : Plu
Registros : 296.022 Filas

Dado que no todas las variables bajo análisis se han guardado en los archivos históricos, en particular los correspondientes a inventario, las variables que son factibles de despejar utilizando otras se procesan y generan, como también los vacíos que presentan algunos períodos.

7.4. Análisis Exploratorio preliminar.

Generada una muestra representativa de los datos, se esta en condicione de realizar un análisis exploratorio preliminar, el proceso se inicia aplicando técnicas de estadística básica, con los resultados que se muestran a continuación.

Característica de las tiendas:

Para efectos de este análisis exploratorio fueron seleccionadas tiendas representativas de cada segmento por tamaño y ubicación geográfica, como lo indica la siguiente tabla:

Ubicación	Grande	Mediana	Pequeña
Norte	1		
Centro Norte		1	1
Centro		2	3
Centro Sur		1	
Sur	1		

.. **Tabla 7. 3 Tiendas por tamaño y ubicación**

Análisis preliminar del Comportamiento de la venta

A continuación se presenta en forma gráfica el comportamiento de la prenda en los distintos tipos de tiendas bajo estudio, que representa lo siguiente:

Las líneas según su color significan lo siguiente:

- Roja representa las ventas
- Amarilla la reposición de inventario
- Verde representa el nivel óptimo de artículos en exhibición (Kanban)
- Zona ploma representa el inventario en tienda (stock)

Tiendas Grandes

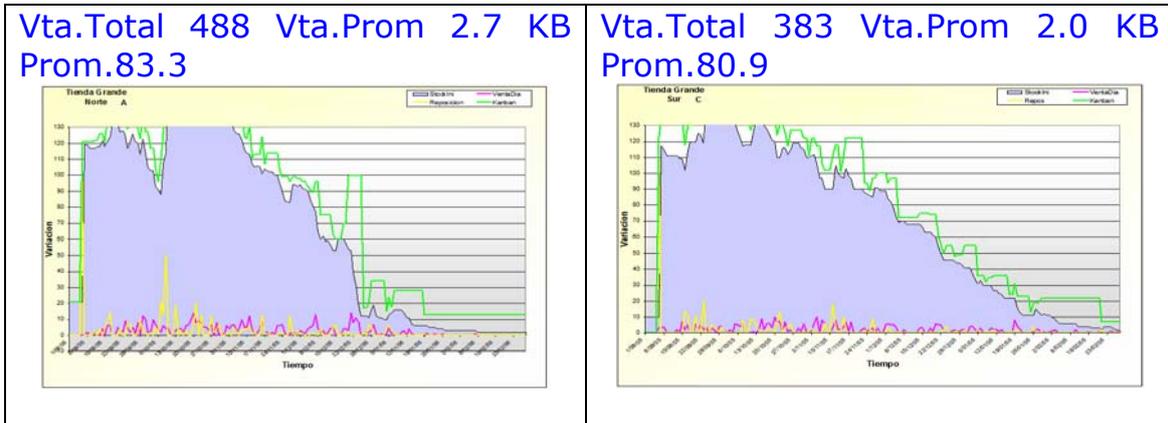


Fig. 7.6 Venta Total en Tiendas Grandes

Tiendas Medianas

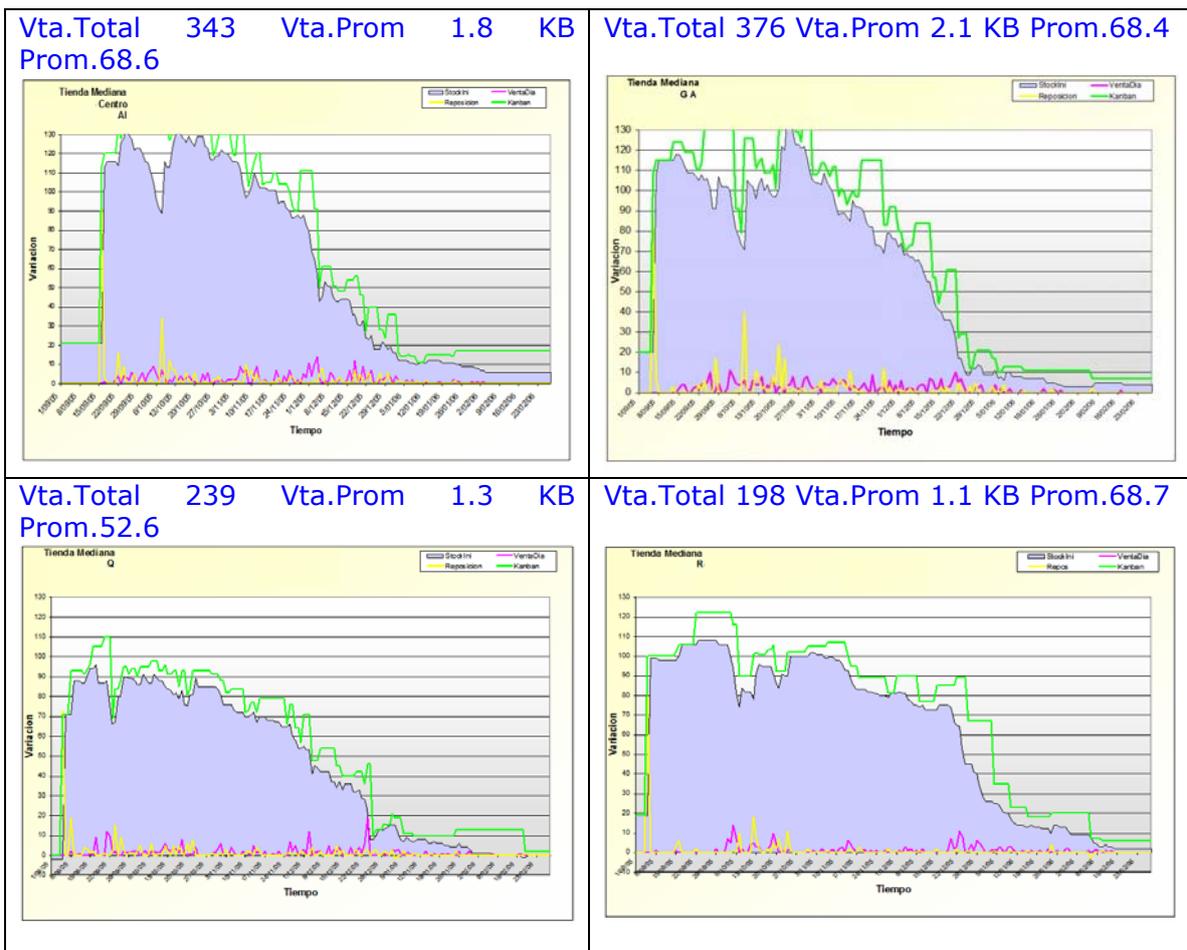
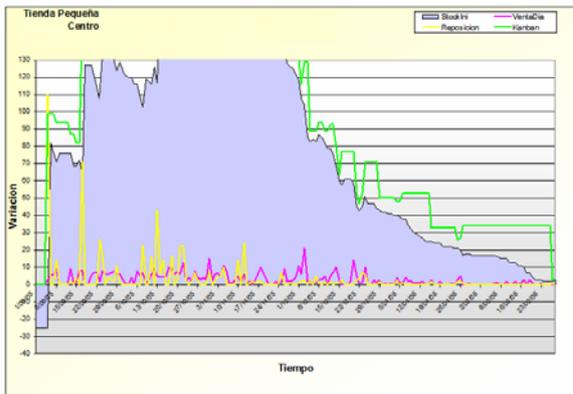


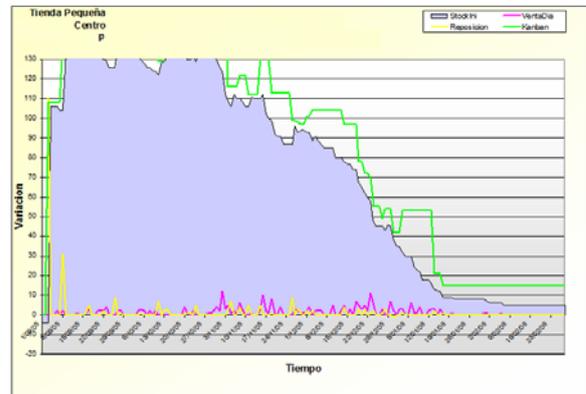
Fig. 7.7 Venta Total en Tiendas Medianas

Tiendas Pequeñas

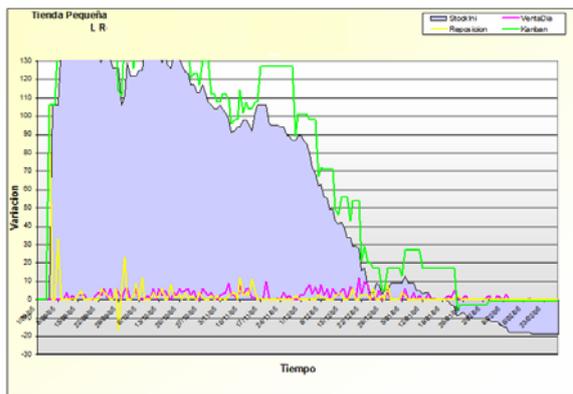
Vta.Total 558 Vta.Prom 3.1 KB
Prom.100



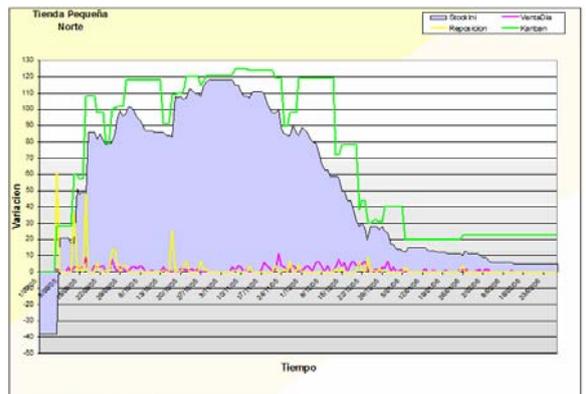
Vta.Total 231 Vta.Prom 1.2 KB
Prom.86.7



Vta.Total 370 Vta.Prom 2.0 KB
Prom.72.7



Vta.Total 255 Vta.Prom 1.4 KB
Prom.70.1



. Fig. 7. 8 Venta Total en Tiendas Pequeñas

Para el análisis exploratorio preliminar se consideran los datos históricos de los últimos tres años donde se verifican que existen estacionalidades como se muestra en la siguiente figura, siendo la venta más relevante la que corresponde al período previo a navidad, en el cual se comercializa más del 20% de la venta total anual. Además existen otras fechas puntuales que dan origen a periodos menos extensos con ventas relevantes como son: en el mes de mayo, el día de la madre y en el mes de agosto el día del niño y también se aprecia un incremento significativo en el periodo de ventas previas a fiestas patrias.

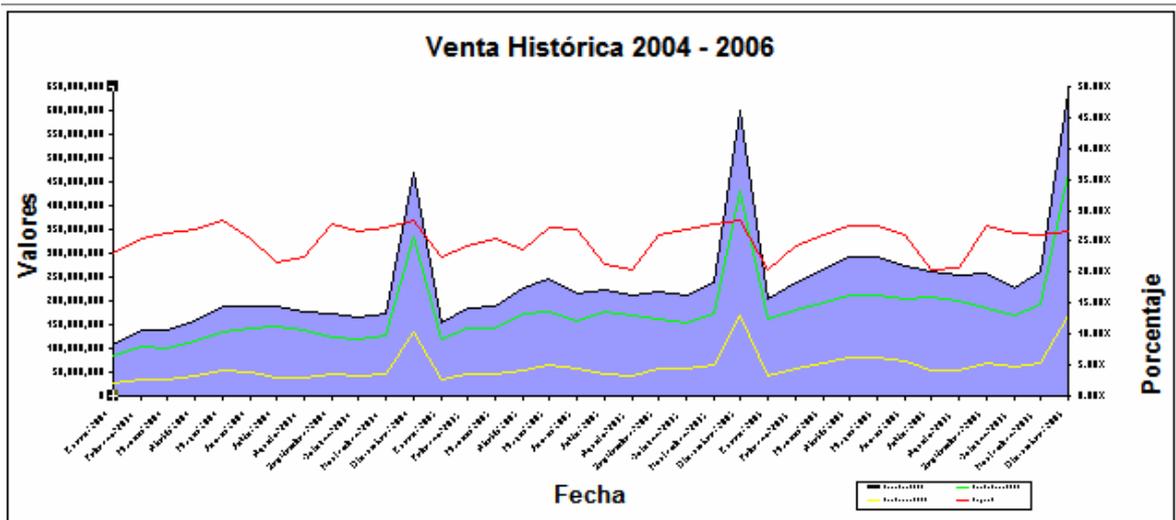


Fig. 7. 9 Venta Histórica 2004-2006

Esta venta histórica está distribuida las 10 líneas de negocios antes mencionadas, estas cuentan con las siguientes participaciones de las ventas:

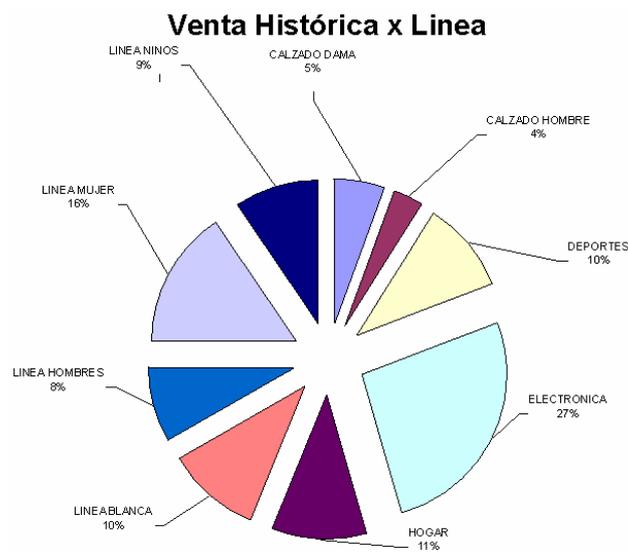


Fig. 7. 10 Venta Histórica x Línea.

En esta distribución de ventas por líneas de negocios se aprecia que la línea electrónica es la que ocupa el primer lugar en volumen de ventas.

Sin embargo, esta línea también ocupa el primer lugar en costos, como se muestra en el siguiente diagrama:

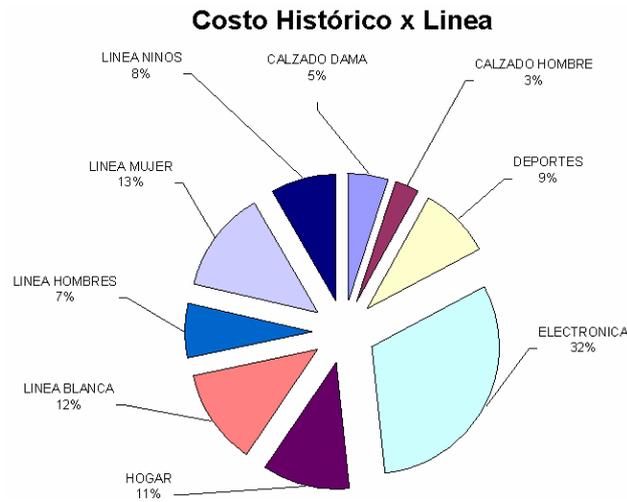


Fig. 7. 11 Costo Histórico por Línea

Esto indica que a pesar del significativo volumen de ventas, no es la línea más rentable, y como se aprecia en el siguiente diagrama, la más rentable corresponde a la línea Mujer.

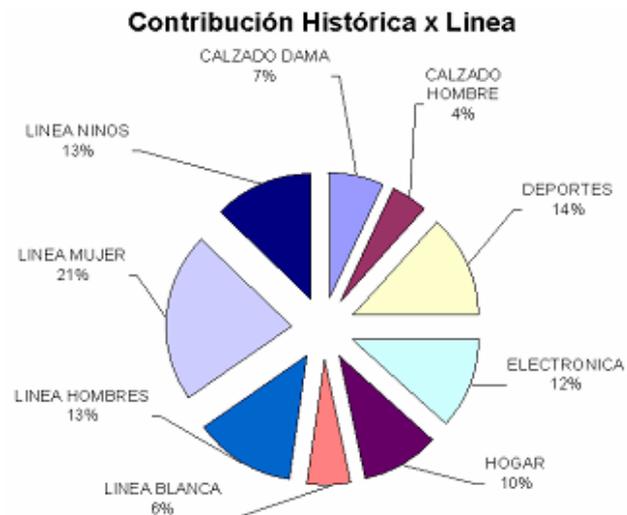


Fig. 7. 12 Contribución Histórica por Línea.

De la interpretación del análisis global del comportamiento de las ventas, los costos y la contribución se destaca lo siguiente:

- Las líneas blandas (vestuario y calzado son mas rentables que las líneas duras (electrónica, electro-hogar, deco-hogar)
- Las líneas duras generan un mayor flujo de dinero, sin embargo, demandan una mayor inversión, generan poca contribución y requieren de una logística y gestión de exhibición distinta.

- También se observa una clara tendencia al alza y un comportamiento similar en los tres períodos analizados
- La tendencia y el comportamiento de la distribución de las ventas, costos y contribución en cada una de líneas de negocios es similar en cada uno de los periodos analizados.

A partir de esta interpretación de los resultados del análisis global, para realizar un análisis más específico se selecciona la línea con mejor resultado para estudiar su comportamiento en el actual periodo primavera. Esta corresponde a la línea de negocios Mujer.

En el análisis del período y línea específica, la venta periódica más alta ocurre los días sábado de esta línea, se revisan el comportamiento global y se observa el mismo comportamiento lo que indica que es una tendencia general de la cadena. El segundo día de mejor venta es el día lunes, con una participación más destacada en la venta del día lunes las tiendas que se encuentran en el centro de la Región Metropolitana.

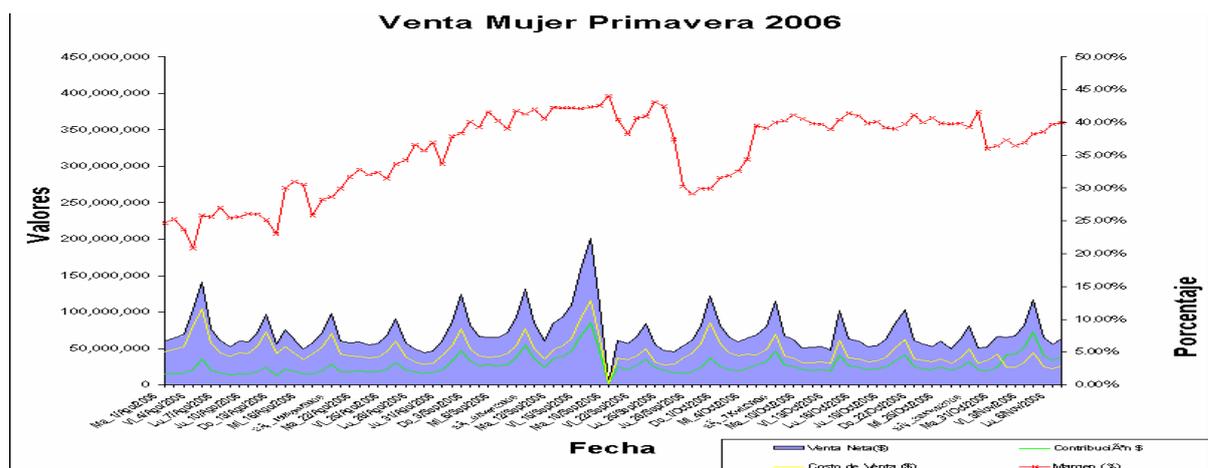


Fig. 7. 13 Venta Mujer Primavera.

La línea Mujer bajo estudio, clasifica sus productos en 11 áreas y distribución de las ventas por área es la siguiente:

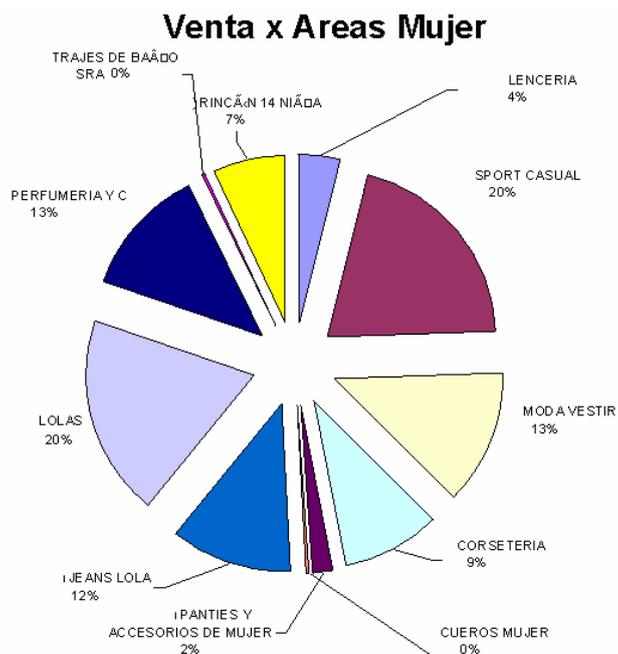


Fig. 7. 14 Venta por Áreas Mujer.

Destaca en esta gráfica las áreas Lolos (Joven Adolescente) y Sport Casual con un 20% de participación de la venta.

Los costos en esta línea de negocios se distribuyen de la siguiente forma:

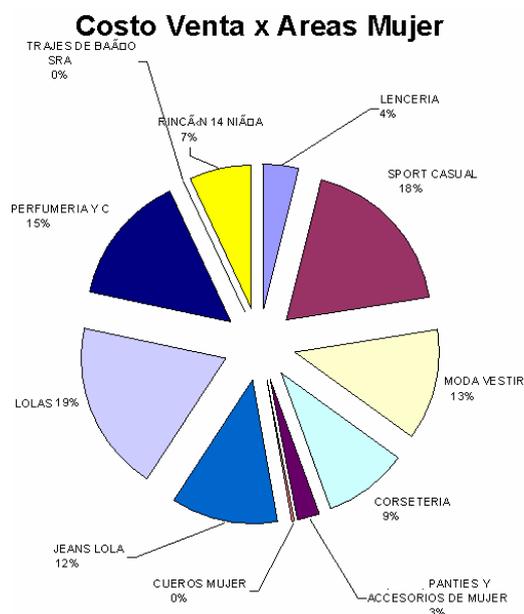


Fig. 7. 15 Costo Venta por Áreas Mujer.

En la distribución de los costos por Área de la Línea Mujer se aprecia que los costos son liderados proporcionalmente por el área Lolos con un 19%.

Esto permite que la contribución de cada una de las áreas de esta línea de negocios se distribuya de la siguiente forma:

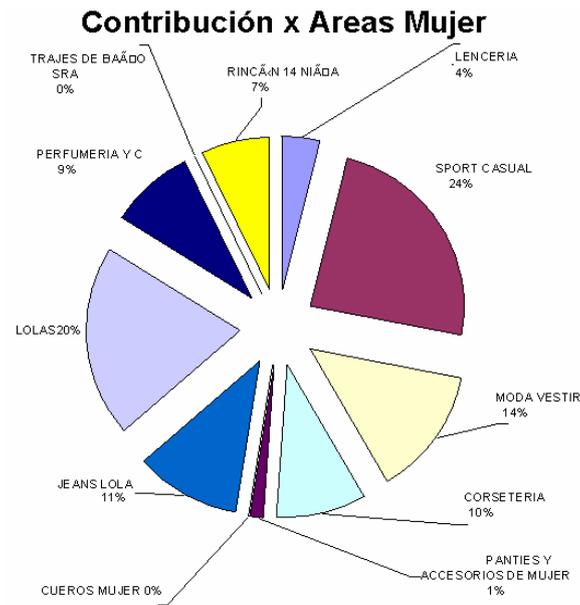


Fig. 7. 16 Contribución por Áreas Mujer.

En la distribución de la Contribución por Áreas de la Línea Mujer destaca la contribución del área Sport Casual.

El área Sport Casual para el período bajo estudio presenta el siguiente comportamiento de compra:

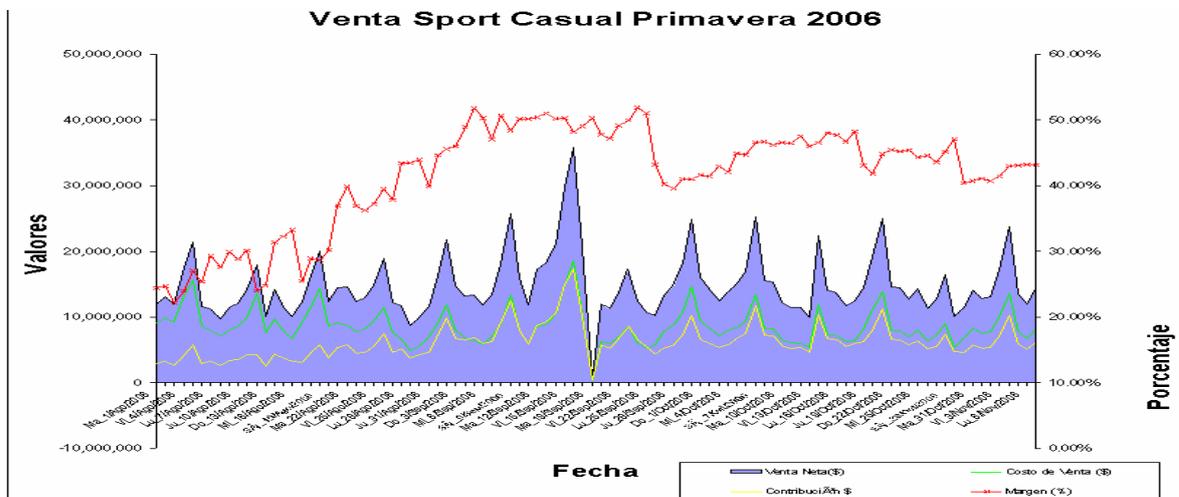


Fig. 7. 17 Venta Sport Casual Primavera 2006.

Las ventas del área Sport Casual se distribuye para el período bajo estudio en 13 prendas, cada una de ellas tiene la siguiente participación de ventas dentro del área Sport Casual:

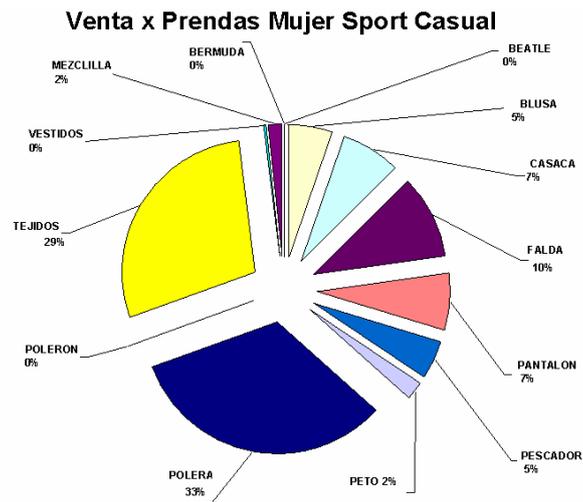


Fig. 7. 18 Ventas por Prendas Mujer Sport Casual.

En las ventas por prendas del área Sport Casual de la línea Mujer, la prenda que realiza en esta temporada la contribución más alta es polera con un 41 % como lo señala la figura siguiente:

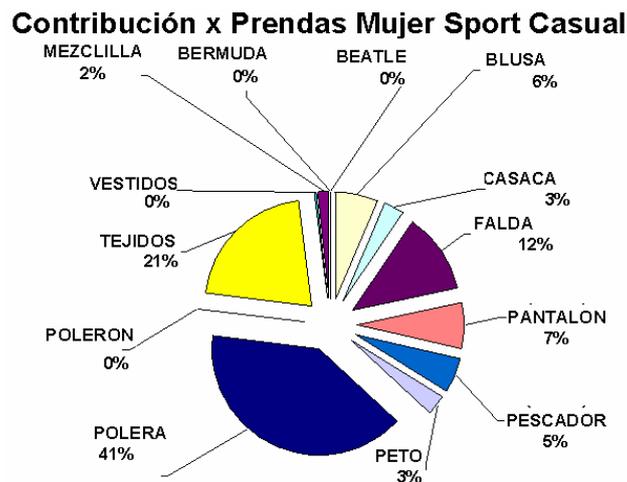


Fig. 7. 19 Contribución por prendas Mujer Sport Casual.

Al analizar el comportamiento de la venta anual de la prenda Polera del área Sport Casual de la Línea Mujer, se puede apreciar que en términos generales tiene un perfil muy similar al de la venta global de la cadena, situación que se verifica en la grafica siguiente:

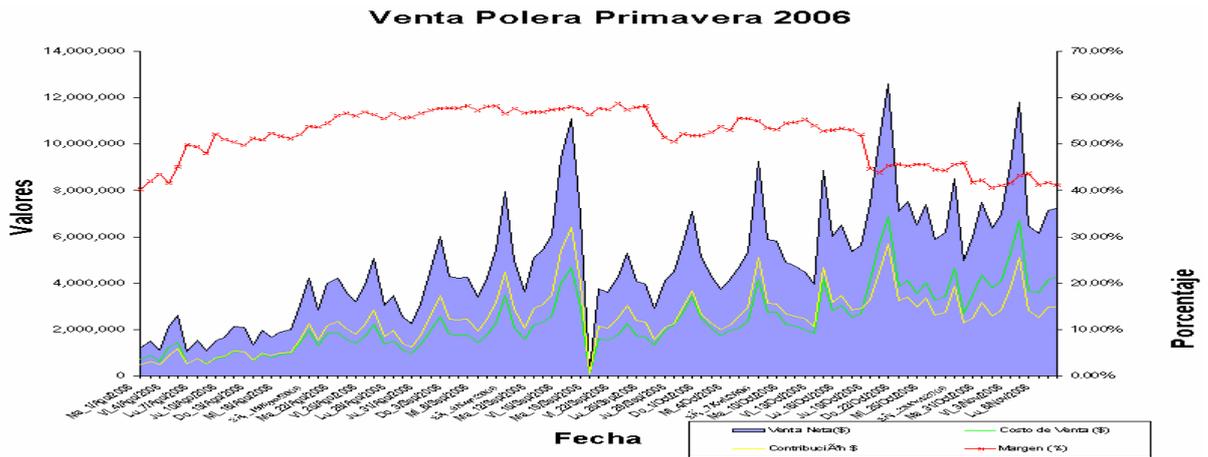


Fig. 7. 20 Venta Polera Primavera 2006.

Al descomponer esta venta del período primavera 2006, en el siguiente nivel de jeraquización de productos, es decir, el nivel de subprenda de poleras, se aprecia que para esta temporada la distribución de la venta por subprenda es la siguiente:

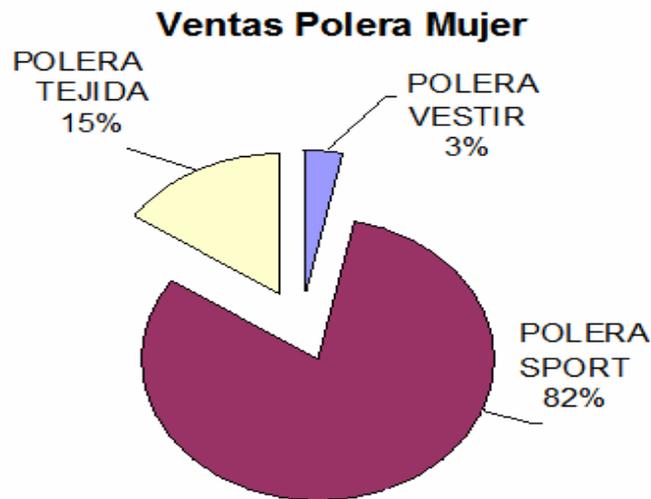


Fig. 7. 21 Ventas Polera Mujer

Sin Embargo, esta misma distribución de la venta para la misma prenda polera por subprenda en el mismo periodo del año anterior, tiene un comportamiento distinto como se aprecia en la figura siguiente:

Contribución Polera Sport Primavera 2005

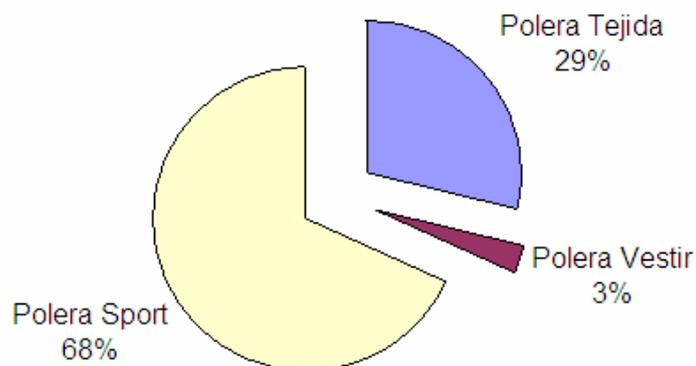


Fig. 7. 22 Contribución Polera Sport Primavera 2005.

Esto demuestra lo drástico que son los cambios de tendencias de una temporada a otra, a nivel de subprenda, el cual en este caso en particular, representa un incremento del 24%, aumento que si no se está al tanto de las nuevas tendencias puede provocar un quiebre de existencia que es muy difícil de revertir por los proveedores dentro de la temporada.

7.5. Interpretar los resultados.

Concluido el análisis exploratorio preliminar, se procede a comparar e interpretar los resultados con las reglas de negocios y los paradigmas que utilizan los especialistas para la toma de decisiones, con los siguientes resultados.

Cuadro Resumen

La siguiente tabla resume las cifras anteriores por tipo de tienda:

Zona	Tienda	Vta. Total	Kb Prom	Vta. Prom
Centro Ah	Pequeña	558	100	3.1
Centro (P)	Pequeña	231	86.7	1.2
Centro (Lr)	Pequeña	370	72.7	2
Norte (C)	Pequeña	255	70.1	1.4
Centro (AI)	Mediana	343	68.6	1.8
Centro (GA)	Mediana	376	68.4	2.1
C. Norte (Q)	Mediana	239	52.6	1.3
C. Sur (R)	Mediana	198	68.7	1.1
Norte (A)	Grande	488	83.3	2.7
Sur (C)	Grande	383	80.9	2.0

Tabla 7. 4 Ventas por tipo de Tienda.

La gráfica asociada a la tabla anterior es la siguiente:

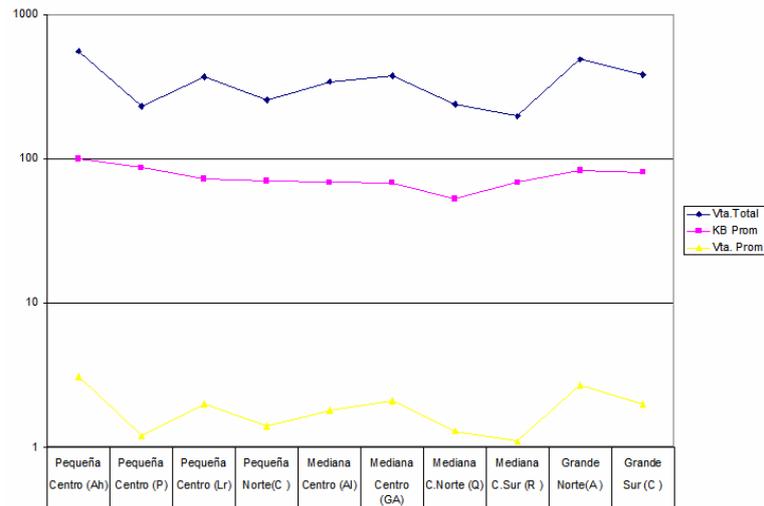


Fig. 7. 23 Ventas por tipo de tienda.

Esta gráfica da cuenta del comportamiento de compra de la muestra, y presenta situaciones como las siguientes:

- La tienda con mayor venta corresponde a una tienda pequeña ubicada en la región metropolitana, es una tienda catalogada por los especialistas como tienda ganadora. Le sigue una tienda grande en la Zona Norte y la tercera en el ranking corresponde a una tienda en la Zona Sur, esto significa que la cantidad vendida no es dependiente ni del tamaño ni de la geo-ubicación de la tienda.
- En las gráficas de detalle por tienda no se aprecia una correspondencia directa entre la cantidad exhibida y la mayor venta producida, en consecuencia se requieren determinar las otras variables que inciden en este resultado.

Con los antecedentes aportados por este análisis exploratorio preliminar y la interpretación de los resultados en conjunto con los expertos en gestión de negocios, se concluye lo siguiente:

- Cada línea de negocios presenta un comportamiento distinto para cada una de sus áreas y prendas.
- Las ventas presentan una distribución diferente dependiendo de las características de tamaño, y ubicación de las tiendas.
- Las ventas presentan distintas distribuciones dependiendo de los días de la semana y de la ubicación de la semana dentro del mes.
- Las ventas son sensibles a los días no hábiles, más aún cuando estos días no coinciden con un fin de semana largo.
- Las ventas son sensibles a los períodos de pago de mensualidades de los tarjeta-habientes, sobre todo aquellos que liberan cupo con la mensualidad y pueden continuar comprando.

- Las ventas son sensibles a las características climáticas y demográficas de las áreas geográficas donde están ubicadas las tiendas.
- Las ventas también son sensibles a la actividad económica que realizan las regiones donde se encuentran ubicadas las tiendas (agricultura; minería, etc).
- Las ventas son sensibles a la tendencia, la moda, la estacionalidad y los cambios climáticos (por ejemplo la venta en años secos con bajas lluvias es diferente a una venta en año lluvioso).
- Las ventas son sensibles a los eventos compañía (evento aniversario) y a los eventos de la industria (día de la madre, día del niño, etc).
- Las ventas son sensibles a los precios, la mezcla de productos y el surtido de los mismos (talla-color).
- Las ventas son sensibles a la ubicación y exhibición de los productos en la sala de venta.

7.6. Análisis comprobatorio (Parte 1: Generar Reglas)

Siguiendo la metodología propuesta, al igual que en el caso anterior, corresponde realizar un Análisis a los métodos, mejores prácticas y herramientas de última generación aplicables al caso bajo estudio.

Este caso, también cuenta con características imprecisas, por tanto, es un buen ejemplo de dominios inciertos, en donde se puede dar lugar a diversas fuentes de incertidumbre, posee un uso habitual del lenguaje natural que presenta información esencialmente implícita; presentándose la multiplicidad, la ambigüedad y la imprecisión del lenguaje natural, ante esta situación, se considera fundamental incorporar un proceso que trate la incertidumbre, lo cual determina la aplicación de las reglas de asociación difusa, para verificar la vigencia del conocimiento del negocio aplicado a este tipo de decisiones de carácter estratégico para la empresa. El carácter estratégico está dado por el impacto que causa una decisión de exhibir o no exhibir una determinada prenda, promover o no un determinado artículo, establecer una baja o no del precio o una determinada promoción.

Tal como ya fue expuesto en el caso anterior, la aplicación de la técnica de extracción de reglas de asociación difusa para este análisis comprobatorio, requiere desarrollar los siguientes dos procesos:

1) Compatibilizar los datos con las variables y contratos usados por los especialistas, elaborando o definiendo el conjunto de etiquetas lingüísticas que permiten hacer imprecisas las variables difusas de las transacciones bajo estudio, para incorporarlas a la base de datos.

2) Posteriormente se desarrolla el proceso de extracción de reglas de asociación difusas, procesando la base de datos, esta vez, con las variables difusas incorporadas en el proceso anterior, para ello se utilizan aplicación de software como FuzzyQuery 2+, de Serrano et al [SJ03] o el software CBA-Aias de Lui [LB99].

7.6.1. Etiquetas lingüísticas

A continuación se describen en orden alfabético los principales datos y cuando poseen características difusas se definen las etiquetas lingüísticas identificadas durante el proceso en estudio:

1. **Área:** este dato describe el primer nivel de agregación de productos perteneciente a una Línea de Negocio a la que pertenece el artículo.
2. **Carga:** es la clasificación que recibe el conjunto de artículos que exhibe un determinado perfil de prendas, esta clasificación depende de varios factores, entre otros: el ciclo de vida del producto, la estacionalidad, los eventos comerciales en proceso.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Liviana:** se asigna esta etiqueta a una Carga inferior al 30% del Perfil.
- **Normal:** se asigna esta etiqueta a una carga alrededor del 80% del Perfil.
- **Pesada:** se asigna esta etiqueta a una Carga que supere el 100% del Perfil, esto significa que la carga desbordará al contenedor, esto se hace cuando se desea promocionar de una manera especial la prenda, por ejemplo, "lleve dos por el precio de uno".

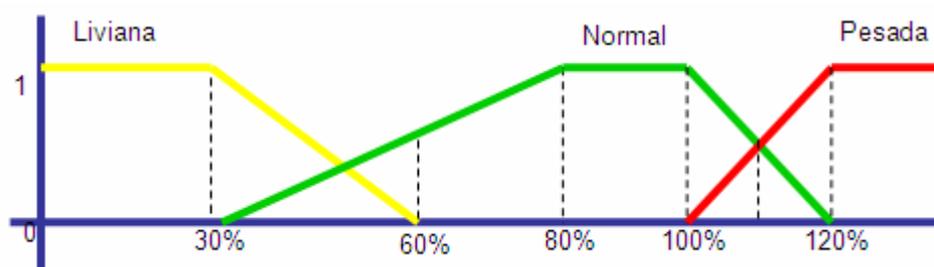


Fig. 7. 24 Etiquetas Lingüísticas de la Carga de artículos en exhibición.

3. **Ciclo de vida del producto (C) o Estacionalidad o Ventana o Temporada:**

Ciclo Inicial (CI); o Avance Temporada o Estacionalidad Baja o Ventana 0 (EI)
 Ciclo Pleno (CP); o Peak o Plena Temporada o Ventana 1 (EP)
 Ciclo Final (CF); o Retiro o Liquidación Temporada o Ventana 2 (ER)

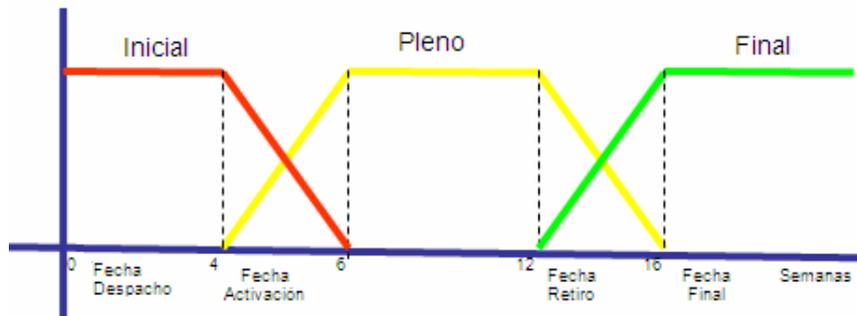


Fig. 7. 25 Etiquetas Lingüísticas del Ciclo de vida del producto

Ciclo de vida: es el periodo de tiempo que transcurre entre la fecha de ingreso o activación y la fecha retiro de tienda que tiene como tope la fecha final. La fecha de retiro se determina tomando en consideración el comportamiento de la demanda y/o la fecha final.

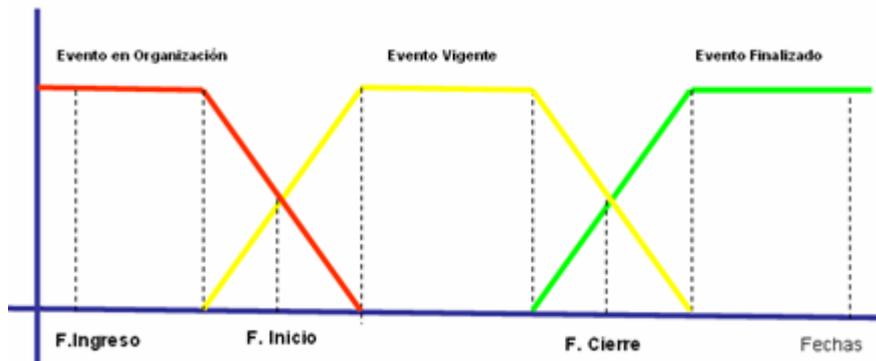
Las etiquetas lingüísticas asociadas a estos atributos son las siguientes:

Ciclo Inicial (CI) o Ingreso: se asigna esta etiqueta al periodo de tiempo que transcurre entre la fecha del primer despacho y la fecha de activación, corresponde al periodo de tiempo que transcurre hasta completar el kanban inicial, cantidad que puede variar dependiendo del comportamiento de la venta y de las acciones de la competencia, entre 4 y 6 semanas (0,0,4,6).

Ciclo Pleno (CP) o Permanencia: se asigna esta etiqueta al periodo de tiempo que transcurre entre la fecha de activación y la fecha final o de traslado o devolución, entre 12 y 16 semanas: (4,6,12,16).

Ciclo Final (CF) o Retiro: se asigna esta etiqueta al periodo de tiempo que transcurre entre la fecha final, o de devolución, o de traslado o retiro, entre 4 y 8 semanas: (12,16,20,20)

4. **Ciclo evento:** es el periodo de tiempo que transcurre entre la fecha de aprobación del evento y la fecha de cierre del evento.



.Fig. 7. 26 Etiquetas Lingüísticas del Ciclo del Evento

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Evento en Organización:** se asigna esta etiqueta al periodo de tiempo que transcurre entre la fecha de ingreso del evento al sistema y su fecha de inicio, corresponde al periodo de tiempo en el cual se coordinan todas las actividades para el éxito del evento comercial.
 - **Evento Vigente:** se asigna esta etiqueta al periodo de tiempo que transcurre entre la fecha de inicio del evento y la fecha de cierre del evento.
 - **Evento Finalizado:** se asigna esta etiqueta a partir de la fecha de cierre del evento.
5. **Clima:** es la tendencia climática que se aprecia para la temporada, que afecta al comportamiento de la prendas y que corresponden a las distintas condiciones propias de cada estación del año (primavera, verano, otoño, invierno).

Las etiquetas lingüísticas se construyen dependiendo la condición climática y su tendencia:

Primavera Fría
 Primavera Lluviosa
 Primavera Cálida
 Primavera Normal o templada
 Verano Frío
 Verano Lluvioso
 Verano Calido
 Otoño Caluroso
 Otoño lluvioso
 Otoño Frío
 Otoño Normal o templado
 Invierno Frío

Invierno Seco
Invierno Lluvioso
Invierno Normal (con lluvias y temperatura baja)

6. **Curva:** corresponde al grupo artículos con una determinada cantidad por cada Talla-Color (PLUs) que lo integran.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Completa:** se asigna al atributo la etiqueta curva completa cuando el stock de todos los PLUs que componen la curva se encuentra por sobre el 20% de su Kanban.
- **Quebrada:** se asigna al atributo la etiqueta curva quebrada cuando la existencia de uno o más PLUs que componen la curva se encuentran bajo el 20% de su Kanban.

7. **Disponibilidad (D) Existencia del Ítems en Bodega:**

Disponibilidad Alta (DA), ítem con sobre existencia.

Disponibilidad Media (DM), ítem con existencias entre el mínimo y alta.

Disponibilidad Baja (DB), ítem con existencias mínima

Disponible Cero (DC), ítem con cero existencia disponible en bodega

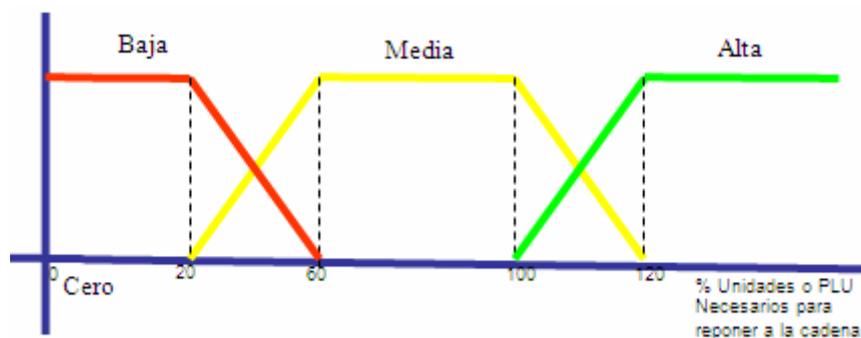


Fig. 7. 27 Etiqueta Lingüística de Disponibilidad.

Disponibilidad (D) Existencia del Ítems en Bodega:

permite conocer la cantidad de artículos existentes en bodega sin distribuir ni asignar. Es la cantidad resultante de restar a la existencia en bodega las cantidades asignadas y distribuidas que aún no se reflejan en los stocks de las tiendas.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Cero (DC):** se asigna esta etiqueta cuando no hay existencia de PLU en bodega (0,0,0,0).
 - **Baja (DB):** se asigna esta etiqueta al nivel de existencias menor al 20% necesario para cubrir la reposición de ítems a toda la cadena (0,0,20,60).
 - **Media (DM):** se asigna esta etiqueta al nivel de existencia que se encuentra entre el 60% y 100% de la cantidad necesaria para realizar las reposiciones a toda la cadena (20,60,100,120).
 - **Alta (DA):** se asigna esta etiqueta a la cantidad de ítems que excede en más de un 120% la cantidad necesaria para cubrir las reposiciones de toda la cadena (100,120,120,120)
7. **Estacionalidad de la Prenda:** se asocia las características climáticas de las distintas estaciones del año para diferenciar las prendas, según las estaciones climáticas del hemisferio sur son las siguientes:

Invierno: período entre el solsticio de invierno y el equinoccio de primavera. Se caracteriza por tener las temperaturas más frías del año, cuando el sol está sobre el hemisferio opuesto. Este periodo va desde el 21 de junio al 22 de septiembre en el Hemisferio Sur, en este periodo se comercializan las prendas denominadas **Prendas de Alto Invierno**, estas prendas se comienzan a exhibir en marzo.

Primavera: período entre el equinoccio de primavera y el solsticio de verano. Se caracteriza por tener las temperaturas cálidas, cuando el sol está sobre este hemisferio sur. Este periodo va desde el 21 de Septiembre al 22 de Diciembre, en este periodo se comercializan las prendas denominadas **Prendas de Verano**, estas prendas se comienzan a exhibir Agosto.

Verano: período entre el solsticio de verano y el equinoccio de otoño. Se caracteriza por tener las temperaturas más cálidas del año, cuando el sol está sobre este hemisferio. Este periodo va desde el 21 de diciembre al 22 de marzo en el Hemisferio Sur, para este periodo se comercializan las prendas denominadas **Prendas de Alto Verano**, estas prendas se comienzan a exhibir en el mes de Noviembre.

Otoño: período entre el equinoccio de otoño y el solsticio de invierno. Se caracteriza por tener las temperaturas templadas,

cuando el sol está sobre este hemisferio norte. Este periodo va desde el 21 de marzo al 22 de junio, en este periodo se comercializan las prendas denominadas **Prendas de invierno**, estas prendas se comienzan a exhibir en Enero.



Fig. 7.28 Etiquetas Lingüísticas de Estacionalidad de la Prenda.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Avance Invierno:** se asigna esta etiqueta a las prendas que se comercializan a partir de Enero, son prendas de media estación con los colores y tendencias predominantes en la próxima temporada de invierno, su comercialización termina a fines de marzo dependiendo de las condiciones climáticas imperantes.
 - **Alto Invierno:** se asigna esta etiqueta a las prendas que se comercializan a partir de Marzo, son prendas de estación con los colores y tendencias predominantes en la temporada, su comercialización termina a fines de agosto dependiendo de las condiciones climáticas imperantes.
 - **Avance Verano:** se asigna esta etiqueta a las prendas que se comercializan a partir de Agosto, son prendas de media estación con los colores y tendencias predominantes en la próxima temporada de verano, su comercialización termina a fines de noviembre dependiendo de las condiciones climáticas imperantes.
 - **Alto Verano:** se asigna esta etiqueta a las prendas que se comercializan a partir de Noviembre, son prendas de media estación con los colores y tendencias predominantes en la temporada, su comercialización termina a fines de Febrero dependiendo de las condiciones climáticas imperantes.
8. **Estado de artículo en bodega:** es la clasificación que recibe el artículo almacenado en bodega, el estado es asignado por el gerente del producto (Product Manager) y se relaciona con la disponibilidad del producto para ser despachado a tiendas.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Activo:** se asigna esta etiqueta a un producto disponible para ser distribuido.
 - **Inactivo:** se asigna esta etiqueta a un producto que se ha dejado de distribuir.
 - **Pendiente:** se asigna esta etiqueta a un producto que no ha llegado a Bodega.
9. **Eventos:** es un indicador que señala si el producto fue vendido por efectos de una actividad promocional o no, se asigna al momento de la venta. Entre las actividades comerciales que promueven las ventas se encuentran por ejemplo la participación del producto en algún tipo de publicidad (Revista, anuncios en Vía Pública, Prensa, TV, Radio, etc); la participación del producto en alguna Promoción (lleve dos o más por el precio de uno, lleve un conjunto (Combo Oferta) de artículos a un precio dado, etc), o a través de un obsequio (por la compra del producto lleve un regalo).

Los valores que este atributo toma son los siguientes:

- **Revista (R):** se asigna esta etiqueta a un producto promocionado por algún tipo de publicidad.
 - **Promoción (P):** se asigna esta etiqueta a un producto que pertenece a algún tipo de Combo oferta.
 - **Obsequio (O):** se asigna esta etiqueta a un producto cuya venta se promociona por medio del obsequio o regalo de otro producto (por la compra de un celular lleve gratis una mochila).
 - **Crédito (C):** se asigna esta etiqueta a un producto cuya venta se promociona por medio de las facilidades de pago, es decir, a través de oferta cuota (lleve este producto en x cuotas a precio contado o con una tasa de interés preferencial comprando con una tarjeta de crédito determinada).
10. **Existencia de ítems en la Tienda o Stock (S)**

Stock Bajo (SB); existencia en tienda menor al 10% del kanban inicial

Stock Medio (SM); existencia en tienda entre el 60% y el 110% del Kanban inicial

Stock Alto (SA); existencia en tienda superior al 110% del Kanban Inicial

Ítem Quebrado (IQ); existencia del PLU en la tienda es igual a cero lo cual causa un quiebre en la curva del producto.

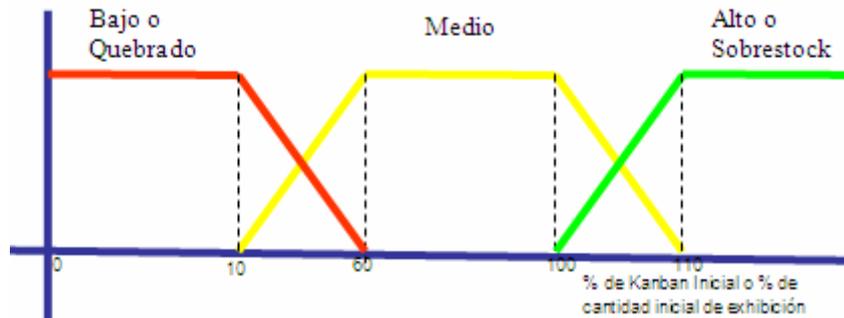


Fig. 7. 29 Etiquetas lingüísticas de Existencia de Ítems en la Tienda (Stock)

Stock: clasifica la existencia de ítems en la tienda en relación a la cantidad de exhibición o kanban del producto.

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Item Quebrado (IQ):** ítem con stock cero (0,0,0,0)
- **Stock Bajo (SB):** se asigna la etiqueta Stock Quebrado a los productos cuya existencia en tienda esta por debajo del 10% del Kanban inicial (0,0,10,60).
- **Stock Medio (SM):** se asigna la etiqueta Stock Medio a los ítems cuya existencia se encuentran por sobre el 60% del kanban inicial y es menor al 110% de la exhibición máxima autorizada (10,60,100,110).
- **Stock Alto (SA) o SobreStock:** se asigna la etiqueta Stock Alto o SobreStock a los productos que exceden por sobre del 110% el máximo de exhibición permitido (100;110,110,110).

11. Item (I) o Producto o PLU (ítem de un determinado modelo talla y color)

Ítem Baja Rotación (IB); o ítem Pegado con rotación menor al 50% de la rotación

Ítem Normal (IN); Producto Normal, con rotación mayor al 50% y menor al 110%

Ítem Ganador (IG); con rotación superior al 110%.

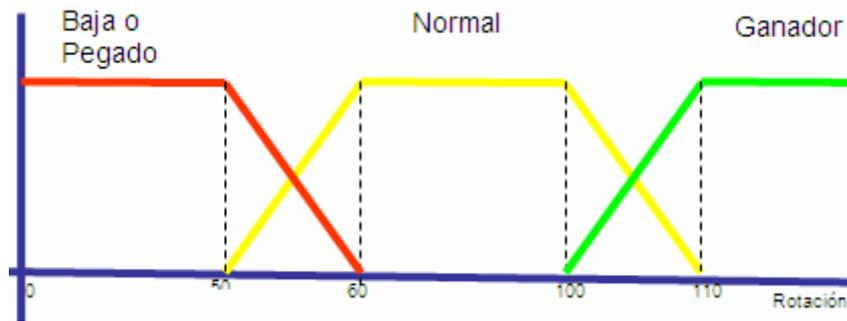


Fig. 7. 30 Etiqueta lingüística para comparar venta diaria (rotación)

Item clasificado comparando la venta con los días venta (Rotación): corresponden a la etiqueta que nos señalan como se comporta la venta en relación a la rotación promedio (días venta), se refiere al número de días promedio necesarios para vender todo el stock. Esta variable se determina para un periodo de venta, habitualmente los últimos 30 días ponderados, por ejemplo, se considera el 20% de la venta promedio ocurrida en los primeros 20 días del mes móvil y el 80% del promedio de la cantidad vendida en los últimos 10 días del mes móvil, esto permite clasificar la venta del producto con las siguientes etiquetas lingüísticas:

- **Item o Producto de Baja Rotación (IB) o Pegado:** se asigna esta etiqueta al volumen ventas que representa una rotación insuficiente para cumplir con la meta de venta (0,0,50,60).
 - **Item o Producto de venta Normal (IN):** se asigna esta etiqueta al producto cuya venta es superior al 60% y hasta 100% de la rotación (50,60,100,110).
 - **Item o Producto Ganador (IG):** se asigna esta etiqueta al producto cuya venta excede la rotación típica (100,110,110,110).
12. **Fecha de fijación del kanban:** este atributo registra la fecha de la última modificación de Kanban para un artículo.
14. **Movimientos de artículos en tienda:** este atributo registra las transacciones que afectan el stock en tienda, acción que se realiza sobre un determinado artículo en una fecha determinada (Venta, Recepción, Traslado, Devolución).

Los valores que este atributo toma son los siguientes:

- **Venta (V):** se asigna esta etiqueta a la transacción que disminuye el stock de un producto en tienda debido a una venta del producto.

- **Recepción (R):** se asigna esta etiqueta a la transacción que aumenta el stock del producto en tienda a consecuencia de un despacho desde bodega o de un traslado desde otra tienda.
 - **Traslado (T):** se asigna esta etiqueta a la transacción que disminuye el stock del producto en tienda a consecuencia de un traslado del producto a otra tienda.
 - **Devolución (D):** se asigna esta etiqueta a la transacción que disminuye el stock del producto en tienda a consecuencia de una devolución del producto a Bodega.
15. **Historial de venta:** este atributo registra el resultado de comparar los días venta de la temporada actual con los días venta del artículo en la temporada anterior, se comparan los días venta del artículo que corresponde a igual día de la fecha del año anterior (primer lunes del mes del año pasado con primer lunes del mes actual, y así sucesivamente).

Los valores que este atributo toma son los siguientes:

- **Sin Historial (-):** se asigna esta etiqueta cuando no se cuenta con datos de la temporada anterior.
 - **Bajo Promedio (B):** se asigna esta etiqueta cuando el resultado de la comparación arroja una cantidad 10% menor que el año anterior.
 - **Promedio (P):** se asigna esta etiqueta cuando el resultado de la comparación se encuentre entre -10% y sobre el 10% de los días venta del año anterior.
 - **Sobre Promedio (S):** se asigna esta etiqueta cuando el resultado de la comparación es un 10% superior a los días venta del año anterior.
16. **Kanban:** este atributo registra la cantidad de PLUs que se desea exhibir en tienda, es el stock deseado en la tienda. Esta cantidad es asignada de acuerdo a la capacidad de exhibición de prendas de cada tienda y es administrado por el gerente de producto para mantener una mezcla óptima de productos.

Kanban (K) Cantidad de PLU para despacho y exhibición en la tienda

Kanban Cero (KC), no esta autorizada la reposición del producto a tiendas.

Kanban Vigente (KV), cantidad > cero, corresponde a la cantidad de exhibición de PLU

Kanban Aumentado (KA), Kanban se ha aumentado con respecto a semana anterior

Kanban Disminuido (KD), Kanban se ha disminuido con respecto a semana anterior

Kanban Sin variación (KS), Kanban sin modificación con respecto a semana anterior

17. **Línea de negocios:** este atributo registra el código del primer nivel de agregación de productos y corresponde a la línea de negocios a la que pertenece el producto.

18. **Margen de contribución:** este atributo asigna una clasificación a los artículos dependiendo del margen de contribución que el producto realiza al negocio.

Los valores que este atributo toma son los siguientes:

- **A:** se asigna esta etiqueta cuando la contribución del producto se encuentra dentro del 20% de productos que aportan el 80% del total de contribución.
- **B:** se asigna esta etiqueta cuando la contribución del producto se encuentra dentro del 20% de los productos que aportan el 80% de los productos excluidos del rango anterior.
- **C:** se asigna esta etiqueta cuando la contribución del producto se encuentra dentro del 20% de los productos que aportan el 80% de los productos excluidos de los rangos anteriores.
- **D:** se asigna esta etiqueta a los productos restantes.

19. **Nivel de rotación:** corresponde a la clasificación por Días Venta.

21. **Perfil:** este atributo registra la capacidad máxima de exhibición de prendas en una tienda (cubicación de tienda). Prenda corresponde al conjunto de artículos de una determinada línea de negocios, por ejemplos Tejidos de Línea Mujer.

22. **Prenda:** este atributo registra el código el tercer nivel de agregación de productos y corresponde a la prenda en el caso de líneas blandas y a sección en el caso de líneas duras a la que pertenece el producto.

23. **Precio (P)**

Precio Rebajado o Precio Oferta (PR), precio de venta menor que precio lista por inicio **de evento, promoción o publicación del ítem en revista.**

Precio Lista (PL), precio de venta igual que el precio de lista (no hay rebaja).

Precio Aumentado (PA), concluye el acontecimiento que dio origen a la rebaja.

24. **Prioridades de reposición:** este atributo asigna una clasificación de prioridades de reposición a los artículos dependiendo si el artículo se encuentra o no participando de un determinado evento o promoción.

Los valores que este atributo toma son los siguientes:

- **Revista (R):** Revista, tiene la primera prioridad de distribución.
- **Prioridad (P):** Prioritario, tiene la segunda prioridad de distribución.
- **Días Venta (D):** el tercer criterio es asignar las siguientes prioridades en orden de días venta por tienda.
- **Producto Nuevo (N):** un producto clasificado como nuevo se distribuye sin restricción de acuerdo al kanban establecido.

23. **Ranking Producto:** corresponde a la clasificación de Días Venta definida anteriormente.

24. **Ranking tienda:** este atributo clasifica a las tiendas de acuerdo a los días venta por prenda o sección.

Tienda (T)

Según nivel de ventas:

⋮

Tienda Ganadora (TG), tiendas con ventas por sobre la venta promedio del producto a nivel de la cadena
 Tienda Regular (TR), tiendas con ventas semejantes a la venta promedio del producto a nivel de la cadena
 Tienda Perdedora (TP), tiendas con ventas por debajo de la venta promedio del producto a nivel de la cadena

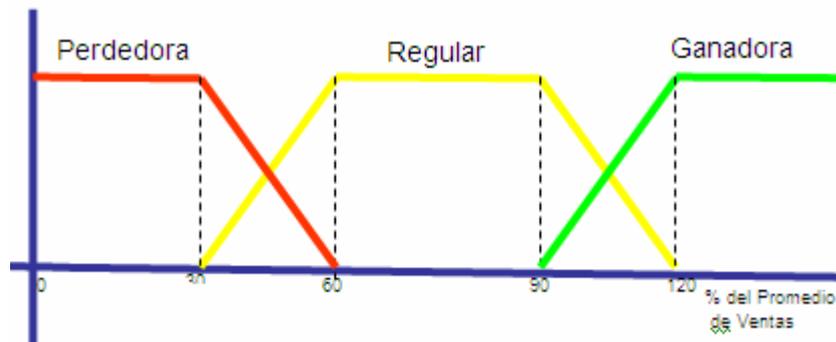


Fig. 7. 31 Etiquetas lingüísticas Según nivel de ventas de la tienda

Jerarquía de Tiendas: las tiendas se clasifican de acuerdo a los días/venta del área o línea de negocios, por tanto una tienda puede ser perdedora para una determinada línea o área de negocios y ganadora para otra.

- **Tienda Perdedora (TP):** se asigna la etiqueta Tienda Perdedora a la tienda que se encuentra por debajo del 30% de la venta promedio del PLU (0,0,30,60).
- **Tienda Regular (TR) o Normal:** se asigna la etiqueta Tienda Regular o Normal a la tienda cuyas ventas están por sobre el 30% y menos del 90% de la venta promedio (30,60,90,120).
- **Tienda Ganadora (TG):** se asigna esta etiqueta a las tiendas que se encuentra por sobre el 120% de la venta promedio (90,120,120,120)

Según Ubicación (U)

Tienda Mall (UM), tienda ubicada en Mall en cualquier región del país

Tienda Regiones (UR); tienda ubicada en regiones y no mall

Tienda Centro (UC): tienda ubicadas en la Región Metropolitana pero no en Mall

Según dimensión o tamaño de la superficie del Local (L)

Tienda con Superficie del Local Grande (LG), tienda que asigna una superficie grande del local para la línea de productos.

Tienda con Superficie del Local Mediana (LM), tienda que asigna una superficie entre pequeña y grande para la línea.
Tienda con Superficie del Local Pequeña (LP), tienda que asigna una superficie del local pequeña para la línea.

25. **Responsable de reposición:** este atributo señala quien es el responsable de reponer el artículo a tienda.

Los valores que este atributo toma son los siguientes:

- **Bodega (B):** es la bodega la responsable de reponer los artículos a tienda.
- **Proveedor (P):** es el proveedor quien es el responsable de reponer artículos a tienda.
- **No reponer (N):** significa que el artículo tiene bloqueada su reposición.

26. **Temporada:** corresponde a la época del año perteneciente al periodo comercial que agrupa las estaciones Otoño-Invierno y Primavera-Verano (ver definiciones en Estacionalidad y tipo de prenda)

Los valores que este atributo toma son los siguientes:

- **Invierno (I):** corresponde al periodo Otoño-Invierno.
- **Verano (V):** corresponde al periodo Primavera-Verano.

27. **Tiendas:** es el código numérico que identifica a cada una de las sucursales de la cadena.

28. **Tipo de prenda:** corresponde a la clasificación de artículos que asigna el gerente de producto al tipo de producto, su definición esta en Estacionalidad y Tipo de Prenda.

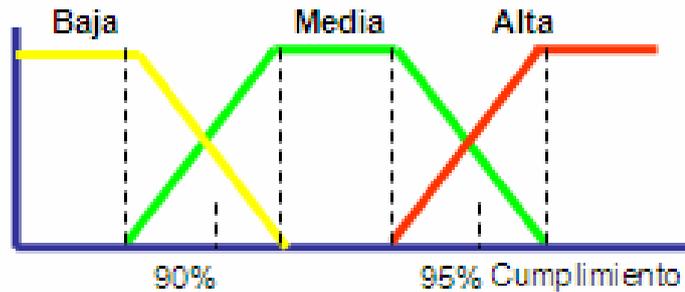
29. **Ubicación:** este atributo señala la zona geográfica a la que pertenece la sucursal, para este efecto el país a sido segmentado en tres regiones: norte, centro, sur.

Los valores que este atributo toma son los siguientes:

- **Norte (N):** corresponde a la zona geográfica comprendida entre la primera y la IV región.
- **Centro (C):** corresponde a la zona geográfica comprendida entre la quinta y la séptima región.

- **Sur (S):** corresponde a la zona geográfica comprendida entre la octava y la XII región.

30. **Velocidad del Operador Logístico:** este atributo clasifica a la velocidad del operador logístico en tres categorías dependiendo de la velocidad que este tenga para despachar el producto a bodega, conjugados con la velocidad y puntualidad de entrega a la tienda.



.Fig. 7.32 Etiquetas Lingüísticas del Velocidad del Operador Logístico

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Baja:** se asigna la etiqueta Operador Logístico Bajo al proveedor que no cumplen sistemáticamente con el 90% con sus fechas de ingreso de productos.
- **Media:** se asigna la etiqueta Operador Logístico Normal, a los proveedores que cumplen sistemáticamente con sus fechas de entrega por entre el 90 y el sobre el 95%.
- **Alta:** se asigna la etiqueta Operador Logístico Alto, a los proveedores que cumplen sistemáticamente con sus fechas de entrega por sobre el 95%.

31. **Velocidad de reposición del proveedor:** este atributo clasifica al proveedor según la velocidad de despacho en tres categorías dependiendo de la velocidad que este tenga para despachar el producto a bodega.

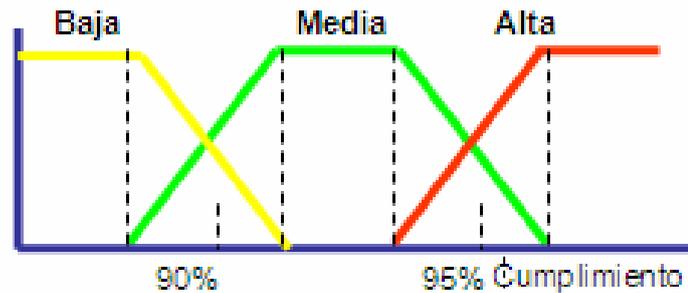


Fig. 7. 33 Etiquetas Lingüísticas Velocidad de reposición del Proveedor

Las etiquetas lingüísticas que este atributo toma son los siguientes:

- **Baja:** se asigna la etiqueta Velocidad de Reposición del Proveedor Bajo al proveedor que no cumplen sistemáticamente con el 90% con sus fechas de entrega de productos.
- **Media:** se asigna la etiqueta Velocidad de entrega del proveedor Media, a los proveedores que cumplen sistemáticamente con sus fechas de entrega por entre el 90 y el sobre el 95%.
- **Alta:** se asigna la etiqueta Operador Logístico Alto, a los proveedores que cumplen sistemáticamente con sus fechas de entrega por sobre el 95%.

32. **Venta esperada:** este atributo clasifica la venta futura del artículo en tres categorías dependiendo de las expectativas generadas en el comportamiento pasado y las condiciones que se espera tener.

33. **Venta (V)**

Venta Mala (VM); venta baja, el monto en pesos de la venta esta por debajo del monto de la Venta Promedio estimada para la temporada (habitualmente se estima en relación al monto de venta promedio de la temporada anterior).

Venta Regular (VR); monto de venta similar al monto de la Venta Promedio estimada para la temporada.

Venta Buena (VB); monto de venta por sobre el monto de la Venta Promedio estimada para la temporada.

Venta Caída (VC); monto de venta menor a la moto de la venta de la misma fecha de la temporada anterior.

Ítem Venta Cero (IC); Producto Pegado o Ítem con venta acumulada = cero

Ítem Perdedor (IP); Producto Perdedor, monto de la venta promedio inferior al monto de la venta promedio de la temporada anterior

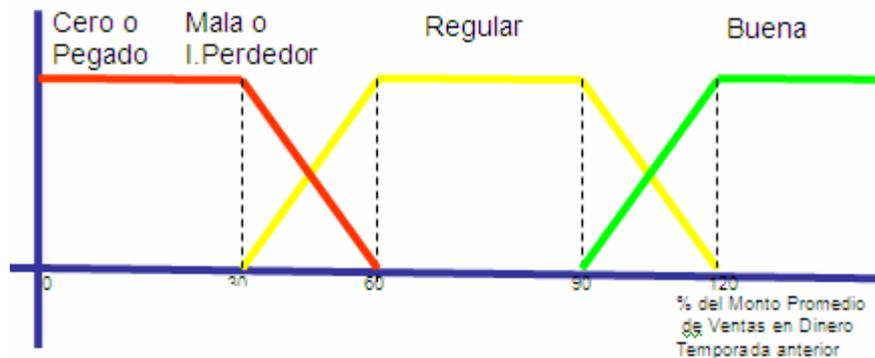


Fig. 7. 34 Etiquetas Lingüísticas de la Venta

Venta Mala (VM); monto de venta inferior al 30% del monto de la venta promedio de la temporada anterior del PLU (0,0,30,60)

Venta Regular (VR); monto de la venta entre el 60% y 90% del monto de la venta promedio de la temporada anterior del PLU (30,60,90,120).

Venta Buena (VB); monto de venta por sobre el 90% del monto de la Venta Promedio de la temporada anterior (90,120,120,120)

7.6.2. Búsqueda de Reglas de Asociación

Los experimentos se realizan con una muestra de las ventas de 11 tiendas durante 20 semanas de un artículo de la línea Mujer con 276 PLUs con la mejor venta y margen de la temporada:

- Los experimentos a nivel de Ventas Diarias x PLU x tienda consideran 85.820 filas.
- Los experimentos a nivel de Venta Semanal x PLU x tienda consideran 16.103 filas.

La muestra de datos utilizados en cada caso presenta las siguientes características:

Ventas	Total Filas	Filas con Ventas	%	Total PLUs	Total Tiendas
X día	85.820	7.319	7%	276	11
xSemana	16.103	6.385	25%	276	11

Tabla 7. 5 Datos utilizados en los Experimentos x día y por semana.

Estas características denotan la gran dificultad que tienen los usuarios para manejar los datos a nivel de PLU, como se aprecia, del

total de Ventas x Día, sólo el 7% de las filas indican ventas. Al agruparlas por semana estas alcanzan al 25% de las filas.

Y la distribución de filas por cantidad de ventas por día y semana respectivamente es la siguiente:

FilasVta=0		FilasVta=1		FilasVta=2		FilasVta=3		FilasVta=4		FilasVta=5		FilasVta=6		FilasVta=7	
Cant.	%	Cant.	%	Cant.	%	Cant.	%	Cant.	%	Cant.	%	Cant.	%	Cant.	%
79415	93	5596	7	734	1%	53	0,01	20	0,002	0	0,000	0	0,000	0	0,000
12013	75	2618	16	908	6%	380	0,02	117	0,073	55	0,034	9	0,001	2	0,000

„Tabla 7. 6 Distribución de filas por cantidad de ventas por día y por semana

Los datos son reales y corresponden a Agosto a Noviembre y es el periodo de venta baja. El periodo de mayor demanda corresponde a Diciembre, esto explica el bajo nivel de venta de la muestra de datos. Este sesgo se tiene presente para el análisis que se desarrolla a continuación.

Para la extracción de las reglas de asociación con el nivel de soporte, confianza y factor de certeza requeridos por el usuario, se utilizan por una parte la aplicación de software Fuzzy Query 2+ de Serrano et al [SJ03], y en paralelo el software CBA y Aias desarrollado por Liu et al [LB99], ambos software utilizan el algoritmo A-priori, con niveles de rendimiento semejantes. A modo de ejemplo se presentan a continuación los registros de algunos procesos realizados con los datos de Ventas por Día y los datos de Ventas por semanas indicando para caso los umbrales, los ítems seleccionados en cada proceso y la cantidad de reglas generadas en cada ejecución:

Ejemplos de Experimentos con Ventas Diarias:

Minsupp	Minconf	MinCertFact	Items	Reglas c/2 Elementos	Reglas C/3 Elementos	Reglas c/4 Elementos
0,001	0,001	1,0	18	1.260	13.322	77.884
0.005	0,005	1,0	18	942	4.012	5.670
0.010	0.010	1,0	18	749	2.218	2.043
0.015	0.015	1.0	18	576	1.380	1.060
0.020	0.020	1.0	18	508	1.005	661
0.025	0.025	1.0	18	417	677	360
0.050	0.050	1.0	18	192	194	46

„Tabla 7. 7 Experimentos con ventas Diarias

Ejemplos de Experimentos con Ventas Semanales:

MinSupp	MinConf	MinFCert	Ítems	Reglas c/2 Elementos	Reglas C/3 Elementos	Reglas c/4 Elementos
0.010	0.010	0.3	15	1.432	9.125	29.575
0.025	0.025	0.6	15	968	5.097	13.756
0.015	0.015	0.3	15	562	3.289	10.287
0.005	0.005	0.3	17	927	7.045	29.447
0.015	0.015	0.6	17	1.116	7.317	26.160
0.015	0.015	0.6	12	294	1.707	5.760

Tabla 7. 8 Experimentos con Ventas Semanales

7.7. Análisis comprobatorio (Parte 2: Depurar las Reglas)

Para esta segunda parte del análisis comprobatorio, se cuentan con las siguientes reglas de asociación difusas extraídas de la base de datos preparadas para este efecto:

1.432 reglas con dos elementos

9.125 reglas con tres elementos

29.575 reglas con 4 elementos.

Para iniciar la depuración de las reglas se revisa en primer lugar el comportamiento de los ítems frecuentes, luego, al igual que en el caso anterior, se aplicaran la medidas de interés objetivas y subjetivas.

7.7.1. Análisis de Item frecuentes

Para la primera revisión de los resultados, se revisan el comportamiento de los ítems frecuentes, que se detallan en la tabla 7.9 del anexo G.

En esta tabla se comparan las cifras de los ítems frecuentes con las expectativas de usuario, a modo ejemplo, comparando los resultados del reporte de ítems frecuentes a nivel de venta diaria en relación al conocimiento de donde se tienen las mejores ventas se obtiene:

¿Mejor Venta según?	Respuesta	Experimento	Soporte
Día de la semana	Sábado	Viernes	15.555453300476074%
Semana del mes	1 y 5	2	23.74909818172455%
Ubicación	Mall	RM	46.27751410007477%
Talla	Sin información	M	27.00769007205963%
Color	Sin información	CRUDO	17.53014624118805%
Tienda	Alameda	Puente	15.46323150396347%

Tabla 7. 9 Comparación de ítems frecuentes con expectativas del usuario

En relación al reporte de ítems frecuentes a nivel semanal se obtiene:

Del total de 20 semanas, la semana con mayor venta es la 104 que corresponde a la cuarta semana del mes de octubre con un 6,7%

Del total de 11 tiendas, son dos las tiendas ganadoras 16 (PCentro) y 17 (Concepción) ambas sobre el 10%

7.7.2. Aplicación de Técnicas de Medidas de Interés a las reglas obtenidas:

Al igual que en el caso anterior, para el proceso de depuración de las reglas extraídas se recurre a las medidas de interés objetivas y subjetivas.

En forma previa, se realiza un Análisis con los especialistas recorriendo la tabla 7.11, 7.12 y 7.13 del anexo G, para validar alguna de las reglas que aplican para tomar decisiones, verificándose alguna de ellas, estos resultados que generan expectativas en los resultados que se obtendrán de este proceso.

En este ambiente expectante se procede a reducir el número de reglas extraídas, seleccionando mediante las medidas indicadas, las reglas útiles e interesantes para someterlas a consideración del usuario.

7.7.2.1. Medidas de Interés Objetivas

Para determinar el conjunto de reglas que pueden ser interesantes para el usuario, se inicia el proceso, aplicando en primer lugar la medida objetiva de clasificación de reglas basadas en la fortaleza de la regla utilizando lo señalado por Berzal et al en [BF01].

Con el propósito de enriquecer el conjunto de reglas a analizar, se aplica en primer lugar el algoritmo de Berzal et al [BF03], para obtener reglas muy fuertes que cumplen con la condición de $A+C > 1$, de las cuales se incluye una muestra representativa en la tabla 7.15 del anexo G.

Luego, aplicando el algoritmo a las primeras mil reglas que cumplen la condición de $A+C < 1$, se obtienen un conjunto de reglas muy fuertes, de las cuales se presentan ejemplos en la tabla 7.16 del anexo G.

De esta forma se obtiene un conjunto de reglas muy fuertes como por ejemplo las siguientes:

Regla	Soporte	Confianza	Fact. Certeza
Si KC, AM -> VM; ¬ VM -> ¬ KC v ¬ AM	47,90% 15,78%	99,53 99,26	0,99 0,98
Si AM, IC -> VM; ¬ VM -> IC v ¬ AM	52,28% 13,97%	99,56 98,12	0,98 0,97
Si VM, DC -> AM; ¬ AM -> ¬ VM, DC	64,59% 27,10%	99,48 98,97	0,98 0,98
Si VM, IC, DC -> AM; ¬ AM -> ¬ VM v IC v DC	49,15% 13,12%	99,48 98,46	0,98 0,98
Si KC, VM -> KS, IC; ¬ IC v LG -> ¬ KC v KS	47,85% 4,94%	99,35 98,75	0,98 0,98

.. **Tabla 7. 10 Ejemplos de reglas muy fuertes**

A las cuales se aplicará el procedimiento propuesto por Liu para determinar reglas útiles e interesantes aplicando medias subjetivas que consideran las expectativas del usuario.

7.7.2.2. Medidas de Interés Subjetivas:

Como ya se ha dicho, no es suficiente aplicar técnicas objetivas para determinar las reglas interesantes para el usuario, pues el interés de una regla depende del dominio de conocimientos del usuario y el momento del tiempo en el cual se realiza. Para analizar entonces las reglas obtenidas del proceso anterior con el conocimiento del usuario, se aplican las técnicas propuestas por Bing Liu et al [LB99] denominadas "User Expectation", las que permiten determinar subjetivamente la inesperabilidad y la utilidad de las reglas para el usuario.

Tal como se planteo en el caso anterior, las técnicas usadas para determinar las medidas de interés subjetivas son del tipo de jerarquías de similitud, que se establecen de la siguiente manera:

- e) Jerarquía según la similitud de los ítems del antecedente de las reglas obtenidas en el proceso de minería de datos con los ítems del antecedente de las reglas definidas por el usuario
- f) Jerarquía según similitud de los ítems del consecuente de las reglas obtenidas en el proceso de minería de datos con los ítems del consecuente de las reglas definidas por el usuario
- g) Jerarquía según similitud de los ítems de ambos lados de las reglas obtenidas en el proceso de minería de datos con los ítems de ambos lados de las reglas definidas por el usuario.
- h) Jerarquía según similitud de conformidad de las reglas obtenidas en el proceso de minería de datos con las reglas definidas por el usuario.

A modo de ejemplo, para este proceso se utiliza el conocimiento del usuario que señala:

"Si la tienda esta en Mall entonces es tienda ganadora": UM -> TG

“Si el stock no es alto entonces la venta es mala”: $\neg SA \rightarrow VM$
 “Si el item esta quebrado entonces el Kanban es cero”: $IC \rightarrow KC$

Y aplicando las Técnicas de “User Expectation” de Liu al conjunto de reglas obtenidas en el proceso de minería de datos se tiene:

a. Resultados del proceso de determinar la jerarquía según la similitud de los antecedentes

Al analizar la presencia de los ítems en el antecedente o atributos en el costado izquierdo de las reglas muy fuertes, se determina, por ejemplo, que la afirmación:

“Si la tienda se encuentra en un Mall entonces es tienda ganadora”,

Esto es $UM \Rightarrow TG$, se invalida al examinar los resultados de la tabla 7.11, que indican que en el costado del antecedente, las reglas con el atributo “Tiendas en Regiones” (UR), tienen mayor presencia en las reglas que el consecuente es “Tienda Ganadora” (TG).

A1	A2	A3	->	C1	C2	C3	Supp	Conf	Fcert
'IC'	'UR'		->	'TG'	'LG'		7,206764	56,447624	0,5155697
'VM'	'UR'		->	'TG'	'LG'		7,0390196	56,033684	0,51096547
'AM'	'UR'		->	'TG'	'LG'		7,0390196	56,033684	0,51096547
'KC'	'UR'		->	'TG'	'LG'		6,610343	57,173546	0,52364403

„Tabla 7. 11 Jerarquía según similitud de antecedente $UM \rightarrow TG$ ”

Por otra parte la afirmación:

“Si el stock no es alto entonces la venta es mala”, es decir, $\neg SA \rightarrow VM$ se confirma, dado que, en la tabla 7.12, las reglas con “Venta Mala” en el consecuente, el atributo “Stock Medio” (SM), es el que tiene mayor presencia, en el antecedente.

A1	A2	A3	->	C1	C2	C3	supp	Conf	Fcert
'AM'	'SM'		->	'VM'			69,694435	99,56512	0,9854714
'AM'	'SM'	'PR'	->	'VM'			41,364296	99,5514	0,98501307
'AM'	'SM'	'LG'	->	'VM'			40,879707	99,48584	0,98282284
'AM'	'SM'	'UM'	->	'VM'			30,49203	100	1
'AM'	'SM'	'PL'	->	'VM'			28,330036	99,58501	0,98613596
'AM'	'SM'	'UC'	->	'VM'			26,64014	98,87013	0,9622534
'AM'	'SM'	'LM'	->	'VM'			17,40806	100	1
'AM'	'SM'	'UR'	->	'VM'			12,562114	100	1
'AM'	'SM'	'LP'	->	'VM'			11,40655	99,189705	0,9729296
'AM'	'SM'	'DA'	->	'VM'			2,7025373	99,08887	0,96956074
'AM'	'SM'	'DB'	->	'VM'			2,3919005	99,2269	0,9741722

„Tabla 7. 12 Jerarquía según similitud del antecedente $\neg SA \rightarrow VM$ ”

b. Resultados del proceso de determinar la jerarquía según la similitud de los consecuentes

Al analizar la presencia de los ítems en los consecuentes o atributo en el costado derecho de las reglas muy fuertes, se determina por ejemplo, lo siguiente:

La afirmación

“Si la tienda esta en Mall entonces es Tienda Ganadora”,

es decir, UM -> TG, según lo observado en la tabla 7.13, se invalida debido a que los atributos del costado derecho con mayor presencia en las reglas con atributos “Ubicación Mall” (UM), tienen en el costado izquierdo de las reglas el atributo “Tienda Regular” (TR), en el costado derecho.

A1	A2	A3	->	C1	C2	C3	Supp	Conf	Fcert
'UM'			->	'TR'			41,662563	100	1
'UM'			->	'TR'	'KS'		41,662563	100	1

„Tabla 7. 13 Jerarquía según similitud de los consecuentes UM -> TG

Por otra parte la afirmación:

“Si el ítem esta quebrado entonces el Kanban es cero”,

Es decir, IC -> KC, presenta en la tabla 7.14 que las reglas con atributos IC tienen en el costado izquierdo KC, confirmando la regla:

A1	A2	A3	->	C1	C2	C3	supp	Conf	Fcert
'IC'			->	'KC'			62,164593	91,890785	0,7853548
'IC'			->	'KC'	'SM'		61,82287	91,38572	0,77402997
'IC'			->	'KC'	'KS'		61,80419	91,3582	0,773418
'IC'			->	'KC'	'KS'	'SM'	61,462425	90,85313	0,7623049
'IC'			->	'KC'	'DC'		58,50524	86,48174	0,673779
'IC'			->	'KC'	'SM'	'DC'	58,188324	86,01331	0,66503507
'IC'			->	'KC'	'KS'	'DC'	58,176018	85,99499	0,66469663

„Tabla 7. 14 Jerarquía según similitud de los consecuentes IC -> KC

c. Resultados del proceso de determinar la jerarquía según la similitud de ambos lados de la regla

Luego al analizar la presencia de los ítems indicados por el usuario, en ambos lados de la proposición, la ocurrencia de ítems en el antecedente y en el consecuente, se tiene por ejemplo:

Para la afirmación:

“Si la tienda esta en Mall entonces es tienda ganadora”,
 Es decir, la regla UM -> TG, presenta en la tabla 7.15, a las tiendas en regiones (UR), con una mayor presencia cuando en su costado derecho existe el atributo “Tienda Ganadora” (TG), lo que invalida esta afirmación.

A1	A2	A3	->	C1	C2	C3	Supp	Conf	Fcert
'IC'	'UR'		->	'TG'	'LG'		7,206764	56,447624	0,5155697
'VM'	'UR'		->	'TG'	'LG'		7,0390196	56,033684	0,51096547
'AM'	'UR'		->	'TG'	'LG'		7,0390196	56,033684	0,51096547
'KC'	'UR'		->	'TG'	'LG'		6,610343	57,173546	0,52364403

..Tabla 7. 15 Jerarquía según similitud de ambos lados de la regla UM -> TG (costado izquierdo)

Al revisar la presencia de estos atributos en el costado derecho se obtiene, según la tabla 7.16, que las tiendas en mall (UM), tiene mayor presencia en el costado derecho el atributo Tienda Regular (TR).

A1	A2	A3	->	C1	C2	C3	supp	Conf	Fcert
'UM'			->	'KS'			41,662563	100	1
'UM'			->	'TR'			41,662563	100	1
'UM'			->	'TR'	'KS'		41,662563	100	1
'UM'			->	'TR'	'KS'	'SM'	41,33327	99,20961	0,9637135
'UM'			->	'TR'	'SM'		41,33327	99,20961	0,96224904
'UM'			->	'TR'	'KS'	'DC'	39,12151	93,90105	0,76863354
'UM'			->	'TR'	'DC'		39,12151	93,90105	0,76137936
'UM'			->	'TR'	'SM'	'DC'	38,82954	93,20012	0,74118966

..Tabla 7. 16 Jerarquía según similitud de ambos lados de la regla UM -> TG (costado derecho)

Luego, se confirma que para esta prenda y en estas circunstancias las tiendas ubicadas en Mall no son del tipo Ganadoras.

d) Resultados del proceso de determinar la jerarquía según la conformidad de las reglas

Finalmente la técnica que completa lo propuesto por Liu et al [LB99], verifica el grado de conformidad de las reglas obtenidas en este proceso con las expectativa del usuario, determinando el grado de conformidad cuantificando si el número de elementos que conforman el antecedente y el consecuente son los mismos que conforman la definición entregada por el usuario, estableciendo el grado de similitud entre ambas.

De este proceso naturalmente las que tienen mayor puntuación son por ejemplo:

“Si el stock no es alto la venta es mala”: $\neg SA \rightarrow VM$

Lo que se puede verificar en la tabla 7.17, donde la siguiente regla presenta un alto nivel de cumplimiento al tener exactamente el mismo número de elementos tanto en el antecedente como en el consecuente, tal como, lo expresa el usuario:

A1	A2	A3	->	C1	C2	C3	supp	Conf	Fcert
'VM'			->	'SM'			70,067116	100	1

Tabla 7. 17 Jerarquía según conformidad de las reglas \neg SA \rightarrow VM

De esta manera se han aplicado las técnicas propuestas por Liu et al denominadas "User Expectation", las cuales permiten de una sencilla manera determinar reglas que cumplen con las expectativas del usuario.

7.8. Incorporar los resultados a la lógica del negocio

Para concluir la aplicación de la metodología propuesta que permite verificar la vigencia y actualizar el conocimiento del negocio, se debe incorporar los resultados obtenidos a la lógica del negocio.

Los resultados obtenidos en este caso, señalan el gran aporte que la aplicación de las reglas de asociación difusa hace para mantener vigente el conocimiento del negocio y contribuir al control de la gestión del mismo.

7.9. Resumen del Proceso de Análisis de la Gestión de Negocios

Este capítulo ha mostrado que la metodología empleada supera los problemas de interés y bondad de las reglas y supera la barrera de comunicación entregando resultados intuitivos para los especialistas.

El empleo de la plataforma computacional en uso también demuestra la superación de las restricciones de rendimiento de los algoritmos.

A modo de resumen de lo obtenido en este caso podemos decir lo siguiente:

El primer gran aporte para áreas de negocios, como el mostrado en este caso, es el convencimiento de la necesidad de implementar tecnología de punta en la toma de decisiones de negocio.

Se ha demostrado la necesidad de conectar la operación diaria para obtener datos desde los procesos transaccionales, a fin de poder realizar una buena tarea de minería de datos para descubrir conocimiento útil e interesante de cómo se comporta el negocio y corregir los paradigmas en uso.

La simple comparación de presencia de los ítems frecuentes con las expectativas del usuario permitió confirmar que el conocimiento del usuario es a nivel global (prenda) y el volumen de datos que generan las transacciones de la compañía dificultan conocer en detalle lo que ocurre a nivel de ítem o PLU, permitiendo identificar paradigmas obsoletos o erróneos.

También se confirma la facilidad que presentan las técnicas elegidas y los beneficios que aportan para la comprensión de los resultados, el lenguaje natural al que el usuario accede, lenguaje que les ayuda a conocer mejor su negocio.

Sorprende al usuario el porcentaje mínimo de PLU que tienen ventas en algunas tiendas, el usuario esperaba una distribución distinta por tiendas. También Esperaba que esta se concentrara en tiendas denominadas a priori por él como tiendas ganadoras.

De igual forma sorprende el comportamiento de la venta en cuanto a los días y semanas de venta, las expectativas del usuario eran claramente distintas.

La aplicación de las técnicas de medidas subjetivas permiten no solo confirmar alguna de las expectativas del usuario, sino también han permitido aportar conocimiento, descartando y reemplazando los paradigmas como por ejemplo que las tiendas en Mall son las tiendas Ganadoras para el caso de la prenda bajo estudio.

Los especialistas del área de negocios reconocen la utilidad de la metodología propuesta y deciden eliminar las reglas obsoletas, actualizar las reglas que han sufrido modificaciones y agregar los nuevos paradigmas a la lógica de negocios existente en los sistemas transaccionales, también decidió que el proceso transaccional almacene aquellos datos que durante este proceso tuvieron que ser generados o deducidos.

Capítulo 8

Conclusiones y Trabajo Futuro

8.1. Conclusiones de los fundamentos teóricos

La investigación bibliográfica y de campo respecto a la madurez informática en las empresas y la importancia de las TI confirman las paradojas características de la Era Digital, existe una abundancia de datos y una carencia de competencias para convertirlos en información, más aún, por estas carencias, no existe uso intensivo de los recursos tecnológicos disponibles para relacionar estos datos con la experiencia acumulada de manera de convertir esa información en conocimiento, conocimiento que constituye la energía que hace la diferencia entre las empresas exitosas y las que no lo son o no pueden sostener las posiciones de liderazgo por mucho tiempo, dado que, llegan a ese estado mediante un enorme esfuerzo de sus áreas especialistas en los temas de competitividad y diferenciación.

La inversión en tecnologías de información sigue centrada en las labores operacionales y las empresas no logran superar la brecha existente entre la ingeniería de negocios y la ingeniería informática, dando origen una incomunicación derivada de la incongruencia entre el lenguaje natural que utiliza la ingeniería de negocios con el lenguaje artificial que utiliza la ingeniería informática tradicional que demuestra su liderazgo en las labores operacionales y su reducido aporte en las labores de gestión estratégica de los negocios.

En el contexto indicado, el objetivo principal de proveer una nueva forma de abordar la toma de decisiones estratégicas, mediante la representación de información imprecisa y técnica de minería de datos, proponiendo una metodología para incorporar esta tecnología informática de última generación a los procesos de toma de decisiones estratégicas en la Gestión de Riesgo y en la Gestión de Negocio en las grandes tiendas de Chile, ha demostrado, por medio de este trabajo, ser plenamente factible de ser realizado.

La Metodología aplicada supera las dificultades de mínimo soporte y confianza, optimiza los tiempos de ejecución y reduce la generación excesiva de reglas y contribuye a hacer más intuitivos los resultados, facilitando la labor de los analistas de gestión estratégica.

8.2. Conclusiones de los casos bajo estudio

Información transaccional dispersa y orientada a las labores operacionales, existen una amplia variedad de sistemas de tratamiento y representación de información que se utilizan en la producción diaria de la empresa, en distintas plataformas de hardware y software, lo cual no facilita la extracción de conocimientos.

El uso de sistemas inteligentes para dar soporte a la toma de decisiones es Incipiente: las técnicas, métodos y herramientas de extracción de conocimiento provenientes del ámbito de la inteligencia artificial se utilizan como elementos de apoyo o elementos complementarios a las soluciones implementadas para resolver las necesidades de la operación diaria, para apoyar la toma de decisiones, se constituye una solución mixta, en general no existe un uso integral de tecnologías inteligentes, para dar solución a una determinada necesidad de conocimiento se recurre a soluciones que mezclan procedimientos manuales, sistemas tradicionales con sistemas o módulos o partes de sistemas inteligentes, con lo cual se confirma, lo que se sostiene en algunas investigaciones, que la extracción de conocimiento depende de cada contexto, por tanto, cada situación requiere un tratamiento ad-hoc, dado que el conocimiento obtenido esta íntimamente relacionado con el contexto al que pertenece la información.

Las decisiones que se toman en el ámbito bajo estudio, son las que se relacionan con dos de los sectores más estratégicos del negocio, las decisiones que afectan a los clientes y a los accionistas. Los encargados de tomar estas decisiones de carácter estratégicos, son personas que aplican el conocimiento común, derivado de experiencia práctica, sus decisiones se basan más en la intuición que en conocimiento extraído de los datos almacenados en las grandes bases de datos que utilizan para realizar las transacciones operacionales diarias.

Los resultados de la aplicación de la Metodología Propuesta permiten señalar que los aportes que realiza son:

1. Reduce la brecha existente entre la gestión estratégica y la informática, permitiendo utilizar los grandes avances en sistemas inteligentes en el ámbito de las decisiones estratégicas de negocios donde es preponderante el uso de la intuición.
2. Provee soluciones más cercanas al lenguaje natural que utilizan los especialistas en gestión de negocios.
3. Automatiza los procesos de determinar la data pertinente, el análisis exploratorio y sistematiza la búsqueda de conocimiento no tácito.
4. Permite a los especialistas concentrarse en el análisis de las nuevas condiciones del negocio y no distraer su tiempo en las labores previas de exploración.
5. Aumenta la disponibilidad de tiempo dedicado al análisis permitiendo ampliar la cantidad de productos bajo observación.
6. Automatiza el proceso de análisis básico dejando en manos de los especialistas las decisiones fundamentales.

7. Los aportes más significativos de la metodología empleada tiene relación con rentabilización del conocimiento y la experiencia de los especialistas mejorando la rentabilidad del negocio.

8.3. Conclusiones Generales

Las principales necesidades de obtención de conocimiento útil desde grandes repositorios, pueden ser resueltas utilizando técnicas de computación flexible, contribuyendo con soluciones eficientes con el nuevo escenario competitivo que demanda acciones rápidas y efectivas de sus analistas de gestión en distintos ámbitos.

Los motivos que han impedido obtener buenos resultados haciendo uso de herramientas inteligentes tienen su origen en lo heterogéneo de los sistemas de tratamiento y representación de la información junto con las imperfecciones que los datos presentan producto del modelo tradicional que sustenta estos sistemas, todo ello constituye las principales dificultades que han impedido obtener buenos resultados, las técnicas de minería de datos relacionadas con la lógica difusa permiten mitigar este riesgo.

La carencia de herramientas tecnológicas para manejar en forma eficiente la imprecisión del lenguaje natural, lenguaje donde priman la información imperfecta y se privilegia el uso de la intuición ofrece un terreno fértil para el uso de las tecnologías provenientes de la computación flexible. Este trabajo permite demostrar que es factible generar una metodología que supere la brecha y utilice el potencial de la computación tradicional y la computación flexible en beneficio de las áreas de toma de decisiones estratégicas.

Finalmente, el fuerte incremento de la complejidad de la gestión comercial producto de la globalización, requiere proveer de herramientas inteligentes que resuelvan en forma más integral la problemática de la extracción de conocimientos desde las bases de datos operacionales para apoyar la toma de decisiones, para ello se deben proveer interfases intuitivas y eficientes como las que el trabajo futuro pretende entregar, y de acuerdo a lo expuesto en este trabajo, la lógica difusa se presenta como una alternativa válida y vigente para tal propósito, sobre todo que la metodología propuesta busca simplificar los resultados, reduciendo la cantidad de asociaciones, descartando mediante algoritmos aquellos poco relevantes y haciendo más intuitivos los resultados a través del uso de etiquetas lingüísticas mas naturales al experto humano, por todo lo anterior, se concluye que es factible otorgar una solución más integral, pro-activa, en línea, para resolver una necesidad de conocimiento para la toma de decisiones comerciales como es el caso

de la prevención y detección del fraude, mediante la extracción de conocimientos utilizando técnicas de lógica difusa aplicadas a las bases de datos operacionales y otras decisiones de carácter estratégico para la organización.

8.4. El trabajo futuro.

El trabajo futuro consiste, por una parte, en extender la aplicación de la metodología propuesta a otras áreas de la empresa y también a otras empresas en otros sectores o ámbitos cuyas decisiones estratégicas de negocios presenten características similares de uso intensivo de la intuición.

Otra parte del trabajo futuro, consiste en el seguimiento a los resultados que se obtienen con el uso de la lógica de negocios modificada a partir de los resultados de los experimentos realizados.

8.5. Síntesis los objetivos de la propuesta de trabajo futuro

Incorporar la metodología Fuzzy Business Intelligence a otros segmentos y líneas de negocios de las empresas bajo estudio.

Aplicar la metodología Fuzzy Business Intelligence a la base histórica de clientes que utiliza el área de Gestión de Riesgo a fin de determinar reglas de comportamiento que permitan obtener una asertividad mayor a la lograda, considerando por ejemplo:

- Lugares frecuentes de compra.
- Líneas de productos de compra habitual.
- Número habitual de compra de productos.
- Frecuencia de compra del mismo producto.

Modelar y proponer un sistema que permita discriminar entre las transacciones genuinas y transacciones ilícitas, utilizando reglas de asociación difusa.

Modelar y proponer un sistema que genere la información útil y oportuna para mantener vigentes los indicadores de gestión que conformen un tablero de control eficiente y proactivo de la gestión del negocio.

Implementar la metodología propuesta en sistemas en línea utilizando los modelos anteriores que al detectar transacciones o condiciones con las características que afecten o perjudique los intereses del negocio y notifiquen a los analistas con antecedentes

suficientemente intuitivos a fin de reducir al mínimo la incertidumbre en la toma de decisión.

Modelar y proponer la implementación de una base de datos relacional difusa que almacene las transacciones históricas y mantenga vigente el conocimiento adquirido mediante un proceso de extracción de conocimiento de las nuevas conductas que afectan al negocio.

8.6. Beneficios esperados del trabajo futuro

El trabajo futuro propuesto permite visualizar entre otros los siguientes beneficios:

Resultados Intuitivos: los intuitivos resultados mostrados, son un aporte al trabajo de los analistas, es un resultado inmediato de la incorporación de cuantificados lingüísticos, los cuales permiten que la extracción de conocimiento se realice en términos más próximos a la forma en que el ser humano resuelve sus problemas.

Extracción de conocimiento útil: la extracción de conocimientos no explícito desde la bases de datos transacciones, permite actualizar el conjunto de reglas de asociación que contribuyen a prevenir y detectar la actividad inconvenientes para el negocio con resultados fácilmente entendibles por los analistas.

Labor Pro activa: la incorporación de etiquetas lingüísticas y sus grados de pertinencia a las bases de datos operacionales, permitirá contar con los datos necesarios para detectar nuevos comportamientos inconvenientes para el negocio y actualizar el proceso de verificación en línea.

Anexos

Anexo A. Minería de Datos

A1. Aproximación a la minería de datos

En nuestros días existe preocupación por descubrir conocimiento en la enorme cantidad de datos que la empresa almacena como resultado de sus transacciones diarias. Este proceso de extracción de conocimiento a partir de datos esta integrado en varias fases: preparación, almacenamiento, extracción (clasificación, agrupamiento, regresión, asociación, etc), evaluación y difusión del conocimiento, este proceso lo ilustra gráficamente Hernández et al [HJ00], de la siguiente forma:

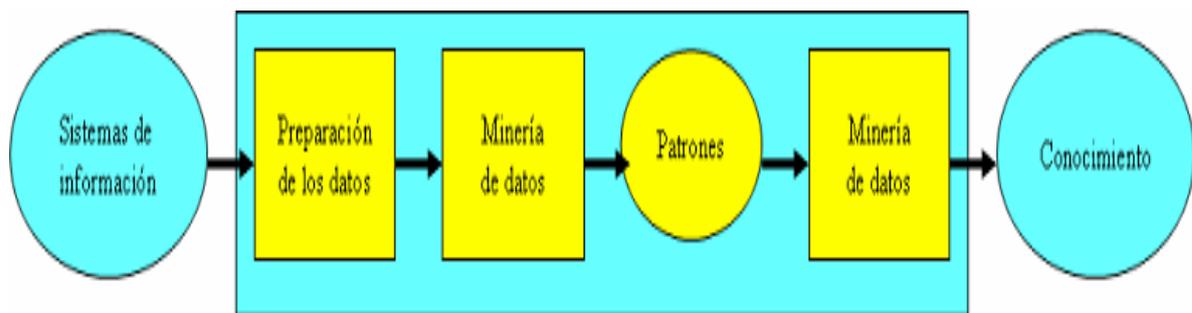


Fig. A.1. Proceso de Extracción de Conocimiento, Fuente [HJ00]

En la definición de Fayyad et al [FU96] se define el KDD como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos”. Definición que considera las siguientes propiedades del conocimiento extraído:

Válido: hace referencia a que los patrones deben seguir siendo precisos para los datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados para su obtención.

Novedoso: que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.

Potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.

Comprensible: la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de

decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de la utilidad).

La minería de datos se relaciona con otras disciplinas, como las que incluye la figura siguiente y que se explican a continuación:



Fig. A.2. Disciplinas que contribuyen a la Minería de datos
Fuente: Introducción a la datos, José Hernández O., Et al, 2000

Bases de datos: fuente de materia prima, son relevantes las técnicas de indización y de acceso eficiente a los datos, estos aspectos inciden en la eficiencia de los algoritmos de la minería de datos.

Recuperación de Información: (información retrieval IR), preocupada de la obtención de información desde datos textuales, es relevante por su relación con la Internet y por su labor de encontrar documentos a partir de palabras claves, aportes relevantes en los procesos de minería de datos que los utilizan como fuente de datos.

Estadística: esta disciplina proporciona muchos conceptos, algoritmos y técnicas a la minería de datos, como por ejemplo, la media, la varianza, las distribuciones, el análisis univariante y multivariante, la regresión lineal y no lineal, la teoría del muestreo, la validación cruzada, la modelización paramétrica y no paramétrica, las técnicas bayesianas, etc.

Aprendizaje automático: al igual que la estadística, esta disciplina aporta algoritmos y programas con sus características propias de contar con la capacidad de aprender un modelo a partir de ejemplos que los usa para resolver problemas.

Sistemas para la toma de decisiones: en la preocupación de proporcionar herramientas informáticas que asistan a los niveles ejecutivos de las empresas e instituciones en la resolución de problemas y en la toma de decisiones, esta disciplina aporta herramientas de análisis como por ejemplo las técnicas ROC (Receiver Operating Characteristic), que provee herramientas que permiten seleccionar el subconjunto de clasificadores que tienen un comportamiento óptimo en general.

Visualización de datos: el uso de técnicas de visualización permiten al observador descubrir, intuir o entender patrones que serían más difíciles de observar a partir de observaciones matemáticas o textuales de los resultados. Técnicas como por ejemplo: graficas de barra, de dispersión, histogramas, cónicas basadas en figuras y colores, jerárquicas entre muchas otras.

Computación paralela, distribuida, en grilla: el aumento de la capacidad de procesamiento y almacenamiento que aporte esta estrategia de uso de recursos permite que la demanda de recursos y el coste computacional de la minería de datos se distribuya en diferentes procesadores y utilice la eficiencia del almacenamiento centralizado en máquinas dedicadas permiten optimizar los algoritmos además de permitir su escalabilidad.

Otras disciplinas: la minería de datos también se nutre de técnicas de otras disciplinas como el lenguaje natural, análisis de imágenes, procesamiento de señales, graficas e imágenes digitales, entre otras, que contribuyen al avance de esta disciplina.

A2. Los modelos de la minería de datos

En la práctica los modelos de minería de datos son dos: el predictivo y el descriptivo.

El modelo predictivo, permite estimar valores futuros o desconocidos de variables de interés que se denominan variable objetivo o dependiente, a través de otras variables conocidas, por ejemplo, estimar la demanda de un producto en función de la promoción realizada. Ejemplos de modelos predictivos son la regresión y la clasificación.

Los modelos descriptivos identifican patrones que explican o resumen datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir datos nuevos. Por ejemplo si se desea conocer las características de un grupo de compradores con el propósito de organizar distintas campañas orientadas a esos grupos objetivos de clientes. Ejemplos de modelos descriptivos son el agrupamiento, las reglas de asociación y el análisis correlacional.

A3. El proceso de Extracción de Conocimiento

El descubrimiento de conocimiento en bases de datos, en inglés Knowledge Discovery from Database, KDD, cuenta de una secuencia iterativa de etapas, fases y procesos, que como lo muestra la figura siguiente, consta de una etapa de rescate de datos, depuración, transformación, minería de datos e interpretación.

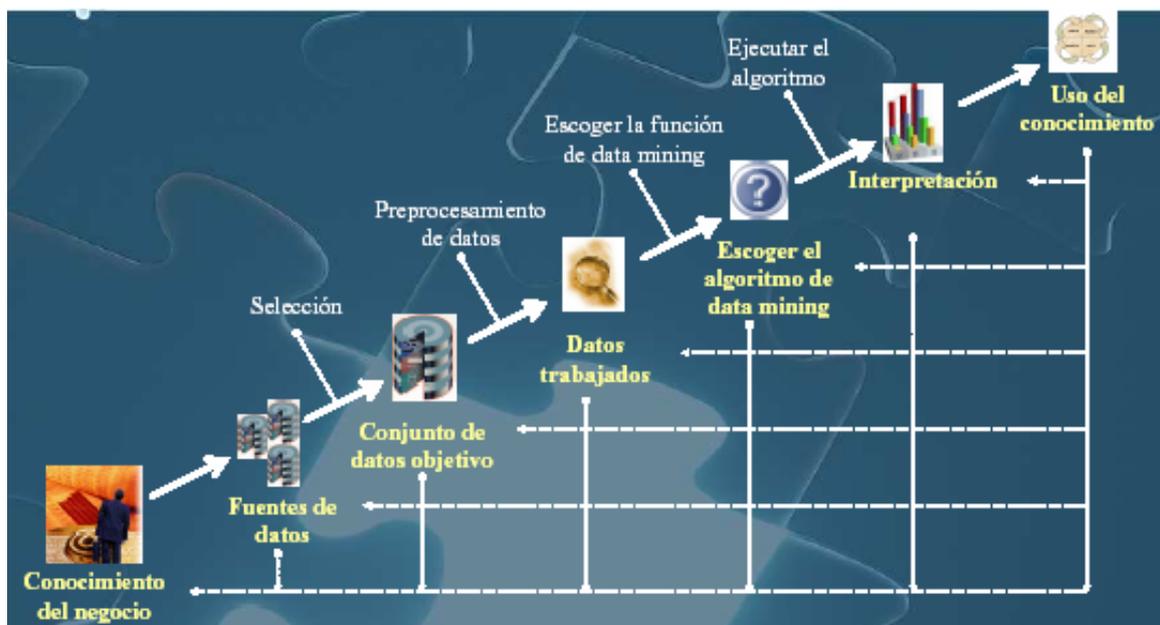


Fig. A3. Etapas y Procesos del KDD, Fuente: Elaborado por Mario José Divan.

En primera etapa o fase se determina la problemática bajo estudio y sus probables las fuentes de información que puedan contribuir a resolver el problema, luego se desarrolla el proceso de obtención de ellos para reunirlos en un repositorio de datos pertinentes.

Dado que los datos proceden de distintas fuentes pueden conducir a información imperfecta, para prevenir esta situación se debe iniciar la segunda fase o etapa que consiste en realizar un proceso de limpieza y depuración, en la que se eliminan o corrigen los datos incorrectos y se complementan los datos incompletos, además de seleccionar las variables o atributos que se consideran relevantes y pertinentes para el estudio a realizar, obteniendo los datos en forma organizada.

A continuación, se procede a transformar los datos, para darles al formato requerido por el proceso de minería de datos, de manera de utilizarlos como elementos de entrada a este proceso.

Estas primeras etapas a las actividades conocidas como preparación de datos y que permiten contar con datos depurados, organizados y en el formato adecuado, iniciándose la etapa o fase de minería de datos (datamining), en la

cual se decide que tarea (clasificar, agrupar, etc) y el método que se va a utilizar.

Posterior y finalmente se da comienzo a la fase de evaluación e interpretación (interpretation), revisando los resultados obtenidos y si es necesario se vuelve a las fases anteriores para una nueva iteración, finaliza el proceso con la fase o etapa de difusión del conocimiento.

A4. Las tareas de la Minería de Datos

Entre las tareas de minería de datos, se encuentran del tipo predictivas y descriptivas que se explican y ejemplifican ha continuación, tomando la definiciones Goebel en [GM00] y los planteamientos de Sánchez en [SD99], el conocimiento a extraer de las bases de datos depende de las necesidades de cada usuario, hay diversos tipos de “patrones” que se pueden obtener a partir de los diferentes algoritmos de extracción, entre los cuales se encuentran los que se describen a continuación y se resumen el la siguiente gráfica:

	Classification	Segmentation	Estimation	Association	Forecasting	Text Analysis	Advanced Data Exploration
Association Rules	●			✓		✓	✓
Clustering	●	✓	●	●		●	✓
Decision Trees	✓	●	✓	✓		●	✓
Linear Regression			✓				●
Logistic Regression	✓		✓			✓	●
Naïve Bayes	✓			●		✓	✓
Neural Nets	✓		✓			✓	●
Sequence Clustering	●	✓	●	●			✓
Time Series					✓		✓

Fig. A.4. Las tareas de la Minería de Datos

Clasificación, se asigna un dato a una clase de entre un conjunto predefinido de clases, para ello, cada instancia (o tupla o registro) cuenta con un atributo denominado Clase de la Instancia, el cual indica mediante su valor a que clase pertenece. Dado un predefinido conjunto categórico de clases, se determina a cual de estas clases pertenece un ítem de dato específico, por ejemplo dadas clases de clientes que corresponden a respuestas a campañas comerciales, identificar la forma de comercialización que más se acerca a un nuevo cliente. Esta es una de las tareas del tipo predictivos más utilizadas. Existen algunas

variantes a esta tarea como son el aprendizaje de ranking, el aprendizaje de preferencias, el aprendizaje de probabilidades.

Regresión, a cada dato se le hace corresponder una variable de predicción real, siendo también una tarea predictiva que consiste en aprender de una función real que asigna a cada instancia un valor real, su principal diferencia respecto a la tarea anterior, es que el valor a predecir es numérico. Dado un conjunto de ítems de datos, la regresión es el análisis de la dependencia de algunos valores de un atributos sobre otros valores de otros atributos en el mismo ítem y la producción automática de un modelo predictivo de estos valores de atributos para nuevos registros, por ejemplo, dado un conjunto de transacciones, construir un modelo que prediga la probabilidad de ocurrencia para nuevas transacciones.

Agrupamiento o Clustering, se agrupan los datos formando clases en función de la similitud entre los mismos, constituye la tarea descriptiva por excelencia y consiste en obtener grupos naturales a partir de los datos. Es distinta a la clasificación, porque a diferencia de esta que analiza datos etiquetados con una clase, Clustering los agrupa para maximizar la similitud entre elementos y al mismo tiempo muy diferente a los objetos de otros grupos. También suele llamarse a esta tarea segmentación ya que divide o segmenta los datos en grupos que pueden ser o no disjuntos. Dado un conjunto de ítems de datos, particiona este conjunto en conjuntos de clases tales que ítems de características similares quedan agrupados, por ejemplo, dado un conjunto de clientes, identificar los clientes que tienen comportamientos de compra similares.

Correlaciones, es una tarea descriptiva que se usa para examinar el grado de similitud de los valores de dos variables numéricas.

Reglas de asociación, también es una tarea del tipo descriptiva, muy similar a las correlaciones que tiene por propósito identificar relaciones no explícitas entre atributos categóricos. Las reglas de asociación no implican una relación causa-efecto, esto significa, que puede no existir causa para que los datos estén asociados. Las reglas se evalúan usando dos parámetros precisión y cobertura (soporte). Las reglas de asociación son una herramienta de minería de datos que permite extraer conocimientos mediante algoritmos que implementan la búsqueda de relaciones entre los atributos utilizando para discriminar el soporte y la confianza, elementos que por si sólo presentan algunos problemas como los señalados en [DS99], [DM99], que son planteados y resueltos por los métodos que propone Sánchez en [DS99], y que por las grandes ventajas respecto de otros métodos que representan estas reglas de asociación para resolver los problemas objeto de este estudio, son revisadas en detalle en el capítulo 2.

Resumen, trata de obtener una representación lo más compacta posible de un conjunto de datos.

Análisis de Secuencias, trata de modelar secuencias de datos, temporales o de otro tipo.

Asociaciones o análisis de enlaces (Link Análisis), dado un conjunto de ítems de datos, se identifican las relaciones entre los atributos e ítems, tales como, que la presencia de un patrón implique la presencia de otro patrón. Estas relaciones pueden ser asociaciones entre atributos con el mismo ítem de datos (por ejemplo alrededor del 64% de los compradores que compran camisas de vestir compran corbatas), o asociaciones entre diferentes ítem de datos (cada vez que cierto stock cae al 5% otro stock se eleva al 13% cada cuatro o seis semanas). La investigación de la relación entre ítems sobre un período de tiempo también se denominan “análisis de patrones secuenciales” (sequential pattern análisis).

Visualización de Modelos: la visualización de modelos juega un rol importante en el descubrimiento de conocimiento entendible e interpretable por humanos, de hecho el sistema ojo-cerebro humano es el mejor reconocedor de patrones que se conoce.

Análisis exploratorio de datos (EDA: Exploratory Data Análisis): EDA es una exploración interactiva del conjunto de datos sin supuestos o modelos preconcebidos de dependencia, a fin de identificar patrones de interés. Representaciones gráficas son usadas para explotar el poder del ojo y la intuición humana.

A5. Las Técnicas o Metodologías de la Minería de Datos

Métodos estadísticos: históricamente el trabajo estadístico ha apuntado principalmente a probar hipótesis preconcebidas y a ajustar modelos a datos. Estas aproximaciones estadísticas usualmente corresponden a un determinado modelo explícito de probabilidad, en general se asume que estos métodos serán usados por estadísticos y la intervención humana es requerida para la generación de modelos e hipótesis candidatas. Estas técnicas no son sólo útiles para regresión, sino que se utilizan para discriminación (clasificación o agrupamiento).

Razonamiento basado en casos (CBR: Case-Based Reasoning): es una tecnología que trata de resolver problemas haciendo uso directo de experiencias y soluciones pasadas. Un caso es usualmente un problema específico que ha sido encontrado y resuelto. Dado un particular nuevo problema, el CBR examina el conjunto de casos almacenados y busca uno de características similares, si el caso existe, su solución es aplicada al nuevo problema, y el problema es agregado a la base de casos para futuras referencias.

Redes Neuronales (NN: Neural Networks): son una clase de sistemas que emulan el cerebro humano, el cerebro humano consta de millones de neuronas que son interconectadas por synapsis, NN son formadas por un gran número de neuronas artificiales conectadas unas a otras de manera similar al cerebro humano. Este paradigma de computación es muy poderoso, permite modelizar

problemas complejos en los que puede haber interacciones no lineales entre variables. Esta técnica se puede usar en tareas de clasificación, de regresión y de agrupamiento. Las redes neuronales trabajan directamente con datos numéricos, los datos categóricos deben numerizarse primero. Las NN tienen una gran capacidad de generalización para problemas no lineales, aunque requieren bastantes datos para entrenamiento.

Árboles de decisión (Decision tree): son una serie de decisiones o condiciones organizadas en forma jerárquica. Los árboles de decisión donde cada uno de los nodos no terminales representa una decisión sobre un ítem de dato, dependiendo de la salida de la prueba, se elige cierta rama. Para clasificar un ítem en particular, se comienza por el nodo raíz y se sigue hacia abajo hasta encontrar el nodo terminal u hoja. Cuando el nodo terminal es encontrado, la clasificación esta hecha. Los árboles de decisión pueden ser interpretados como un conjunto de reglas, caracterizado por su organización jerárquica de reglas. Estas técnicas se usan en tareas de clasificación, agrupamiento y regresión, usados en el primer caso, reciben el nombre de árboles de clasificación y en el último caso reciben el nombre de árbol de regresión.

Reglas de inducción (Rule Induction): la regla establece una correlación estadística entre la ocurrencia de ciertos atributos en el ítem de dato o entre cierto ítem en el conjunto de datos.

Redes Bayesianas (BBN: Bayesian Belief Networks): BBN son representaciones gráficas de distribuciones de probabilidad, derivadas desde el contador de co-conurrencias en el conjunto de datos del ítem de datos. Específicamente una BBN es un grafico directo acíclico, donde los nodos representan atributos de variables y los conectores representan dependencias probabilísticas entre los atributos de las variables. Asociados a cada nodo están las distribuciones de probabilísticas condicionales que describe las relaciones entre el nodo y sus padres.

Algoritmos Genéticos / Programación evolutiva: son estrategias de optimización de algoritmos que son inspiradas por los principios observados en la evolución natural. De una colección de potenciales soluciones a problemas que compete unos con otros, la mejor solución es seleccionada y combinada con otras, esperando obtener la mejor solución del conjunto de soluciones. Los algoritmos evolutivos son métodos de búsqueda en el espacio de soluciones. Dada una población de potenciales soluciones de un problema, la computación evolutiva expande esta población con nuevas y mejores soluciones. Los algoritmos genéticos se pueden usar para el agrupamiento, la clasificación y las reglas de asociación, así como para la selección de atributos.

A6. Los Métodos y Algoritmos

Según el estudio de Goebel [GM00], la distribución de uso no exclusivo de métodos es la siguiente:

Método	Uso
Estadístico	21
Árboles de decisión	18
Redes Neuronales	10
Razonamiento Basado en casos	6
Redes Bayesianas	3
Algoritmos Genéticos	3
Conjuntos ásperos	2
Conjuntos difusos	1

Tabla A.1. Uso de Metodos Estadístico

De acuerdo a esta investigación, los métodos más usados corresponden a los métodos estadísticos combinados con algunas otras técnicas no estadísticas. Luego le siguen los Árboles de decisión cuyos algoritmos de mayor uso corresponde a ID3 y C4.5, a continuación están los métodos de clasificación que utilizan Redes neuronales con algoritmos de retropropagación.

A continuación se revisan los dos métodos no estadísticos de mayor frecuencia de uso en la investigación de Goebel y que además presentan mayor presencia en el ámbito de las soluciones desarrolladas de la detección y prevención de fraude, tal como lo podemos comprobar en sitios dedicados a la investigación, desarrollo y promoción de productos orientados a esta problemática del fraude con tarjetas de crédito, como por ejemplo [SM00],[FB00],[FJ02],[LP02],[BT03].

A6.1. Algoritmo ID3 y C4.5

El algoritmo ID3 (interactive Dichotomizer), fue introducido por Quilan en 1979 y se encuadra dentro de la familia de algoritmos TDIDT (Top-Down Induction of Decisión Tree), a la que también pertenece el C4.5 que constituye la versión definitiva del ID3, la siguiente es la descripción que hace Restrepo [RR00] del algoritmo ID3.

Aunque el algoritmo se desarrolló inicialmente para clasificar movimientos en ajedrez, con atributos categóricos (que podían tomar sólo valores discretos de un conjunto finito) y dos clases, la mayoría de las implementaciones actuales permiten atributos con valores enteros o reales, y clasificación en más de dos clases.

Se somete, como entrada al algoritmo, un conjunto de casos ejemplo en el cual debe haber representantes de todas las clases manejadas por el experto. El conjunto es analizado por el programa, el cual entrega las reglas inducidas.

La representatividad del conjunto de entrenamiento afecta la calidad de las reglas inducidas: el algoritmo es incapaz de descubrir algo que no esté allí. Por esto una buena selección de atributos y ejemplos es fundamental. La selección de atributos debe basarse en el conocimiento acumulado por la experiencia.

Los atributos son los factores que influyen la clasificación o decisión. En este algoritmo cada atributo forma un nodo intermedio en un árbol cuyas hojas

o nodos terminales son las clases o decisiones. Dado el conjunto de ejemplos, el ID3 selecciona el atributo que subdivide los ejemplos *de la mejor manera*. Un atributo categórico con n posible valores subdividiría los ejemplos en n subconjuntos, y será *el mejor atributo* para la próxima regla si cada uno de los n subconjuntos contiene elementos de sólo una clase; sería un mal atributo a seleccionar si los subconjuntos poseen elementos de muchas clases. En el caso de atributos con valores numéricos se requiere escoger valores de corte (umbrales) de manera óptima para la subdivisión, por ejemplo "altura mayor de 51.2" divide los ejemplos en 2 subconjuntos. Como criterio para medir *el grado de mezcla de clases* en los subconjuntos, el ID3 utiliza una *medida de información*, basándose en la teoría estadística de la información de Claude Shannon: selecciona el atributo que produce la mayor disminución de la *información* (en el sentido de Shannon) en los subconjuntos.

Si los atributos se terminan antes de completar el árbol, entonces hay ejemplos contradictorios en el conjunto de entrenamiento, o sea ejemplos con iguales valores de atributos, pero de clases diferentes. El algoritmo básico original no estaba diseñado para esto, y es necesario introducir un atributo extra para distinguir entre ejemplos contradictorios. En otras situaciones el proceso puede revelar carencias en el conjunto de entrenamiento, al identificar un nodo terminal en el árbol inducido, para el cual no hay un ejemplar, o sea que la clase para esa combinación de valores de atributos es desconocida.

La estructura del ID3 es iterativa. Se escoge un subconjunto de los casos de entrenamiento, el cual se denomina *ventana*, y se forma un árbol de decisión a partir de ese subconjunto.

Este árbol clasifica correctamente los casos de la *ventana*. Luego se intenta clasificar con ese árbol los restantes casos del conjunto de entrenamiento. Si el árbol da una correcta clasificación para todos los casos, entonces termina el proceso. Si no, entonces se seleccionan unos de los objetos incorrectamente clasificados, se agregan a la ventana y se repite el proceso, de esta manera, es usual hallar árboles de decisión correctos luego de unas pocas iteraciones para conjuntos de entrenamiento de hasta 30.000 objetos, descritos por medio de hasta 50 atributos.

El árbol de decisión se recorre desde la raíz, y tanto en ella como en cada uno de los demás nodos se decide cuál rama tomar con base en el valor de algún atributo del ejemplar que se esté clasificando, hasta llegar a un nodo terminal (hoja), que, corresponde a la clase en que queda clasificado el ejemplar.

Suponiendo que el conjunto de casos de entrenamiento posee p ejemplares de clase P y n ejemplares de clase N, para el algoritmo ID3 asumimos lo siguiente:

1- Un árbol de decisión correcto clasificará los ejemplares, en cada clase, en las mismas proporciones en que están representados en el conjunto de entrenamiento. O sea que tomando un ejemplar al azar, este se clasificará como P con probabilidad $p/(p + n)$, y como clase N con probabilidad $n/(p + n)$.

2- El árbol de decisión se puede considerar como fuente de un mensaje: " P " o "N", y la información esperada para generar este mensaje será entonces (con logaritmos base 2):

$$I(p,n) = -(p/(p + n))*\text{LOG}(p/(p + n))-(n/(p + n))*\text{LOG}(n/(p + n))$$

Si usamos un atributo A, que tenga v valores posibles, como raíz del árbol, las v ramas salientes de la raíz conducen a su vez a v sub-árboles. Tomando un la i-ésima rama, que tiene pi objetos clase P y ni objetos clase N, la información requerida para ese i-ésimo sub-árbol será I(pi,ni) , y entonces la información esperada requerida para el árbol con A como raíz se obtiene como el promedio ponderado:

$$E(A)=((p_1+n_1)*I(p_1,n_1)+(p_2+n_2)*I(p_2,n_2)+\dots+(p_v+n_v)*I(p_v,n_v))/(p+n)$$

Donde la ponderación para la i-ésima rama es la proporción de los objetos que pertenecen a esa rama, por lo tanto, la información ganada al ramificarse gracias al atributo A es:

$$\text{Ganancia}(A) = I(p,n) - E(A)$$

un buen criterio es escoger el atributo que *gana* la mayor información. ID3 examina todos los atributos y escoge el de máxima ganancia, forma la ramificación y usa el mismo proceso en forma recursiva para formar sub-árboles a partir de los v nodos generados

Para darnos una idea de la aplicación del algoritmo ID3, veamos un ejemplo:

Sea el conjunto de entrenamiento el siguiente, que clasifica el estado del tiempo en dos clases, que denominaremos P y N, según 4 atributos:

Caso#	General	Temper.	Humedad	Viento	Clase
1	asoleado	Caliente	alta	No	N
2	asoleado	Caliente	alta	Si	N
3	nublado	Caliente	alta	No	P
4	lluvioso	Templada	alta	No	P
5	lluvioso	Fría	normal	No	P
6	lluvioso	Fría	normal	Si	N
7	nublado	Fría	normal	Si	P
8	asoleado	Templada	alta	No	N
9	asoleado	Fría	normal	No	P
10	lluvioso	Templada	normal	No	P
11	asoleado	Templada	normal	Si	P
12	nublado	Templada	alta	Si	P
13	nublado	Caliente	normal	No	P
14	lluvioso	Templada	alta	Si	N

Tabla A.2. Conjunto de Entrenamiento del caso Juego de Golf

Nueve objetos son clase P y cinco son clase N, entonces la información requerida para la clasificación es:

$$I(p,n) = - (9/14)*\text{LOG}(9/14) - (5/14)*\text{LOG}(5/14) = 0.940 \text{ bits}$$

Considerando el atributo General, con sus tres valores (v=3):

Para el primer valor, hay 5 objetos que lo tienen, 2 clase P y 3 clase N, entonces:

$$p_1 = 2, n_1 = 3, I(p_1, n_1) = 0.971$$

Analogamente, para el segundo valor posible de A:

$$p_2 = 4, n_2 = 0, I(p_2, n_2) = 0$$

Y para el tercer valor de A: $p_3 = 3, n_3 = 2, I(p_3, n_3) = 0.971$

Por lo tanto el requisito de información esperada, después de chequear este atributo es:

$$E(\text{General}) = (5 \cdot I(p_1, n_1) + 4 \cdot I(p_2, n_2) + 5 \cdot I(p_3, n_3)) / 14$$

$$E(\text{General}) = 0.694$$

Y la Ganancia de este atributo es:

$$\text{Ganancia}(\text{General}) = 0.940 - E(\text{General}) = 0.246$$

Y el mismo procedimiento aplicado a los otros tres atributos da:

$$\text{Ganancia}(\text{Temperatura}) = 0.029$$

$$\text{Ganancia}(\text{Humedad}) = 0.151$$

$$\text{Ganancia}(\text{Viento}) = 0.048$$

Entonces ID3 escoge General como atributo para la raíz del árbol, y se repite el procedimiento para cada uno de los tres sub-árboles. Resulta así que en la rama de "asoleado" se pregunta luego por el atributo "Humedad", en la rama de "nublado" ya se llega a una clasificación como "P" sin necesidad de más chequeos, y en la rama de "lluvioso" se chequea el atributo "viento".

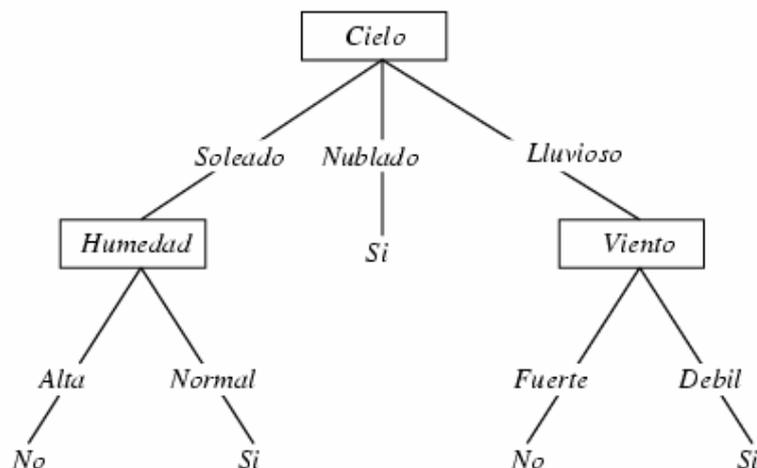


Fig. A.5. Arbol de decisión algoritmo ID3

Anexo B Lógica Difusa

B1. Computación Flexible y Lógica Difusa, una aproximación.

Como se ha planteado en más de una oportunidad durante este trabajo superar uno de los principales desafíos de nuestra época, es vencer la paradoja que somos ricos información pero pobres en conocimiento, otra de las razones a la que se atribuye esta paradoja es el paradigma que utiliza una manera tan distinta al razonamiento humano para representar y organizar los datos, como también las herramientas de manipulación, paradigma denominado computación tradicional o computación rígida o hard computing, que ha sido exitoso para obtener información tácita.

Según los antecedentes que se revisan a continuación, este paradigma fue el resultado de las distintas circunstancias y restricciones existentes como por ejemplo las mínimas capacidades tanto en almacenamiento como en procesamiento existentes en los albores de la computación; la vigencia del modelo de administración de Taylor de la era industrial que privilegiaba la cantidad de operaciones por unidad de tiempo, la carencia de marcos teóricos y conceptuales distintos a la lógica tradicional, entre otros, lo que derivó en el desarrollo y uso de modelos de almacenamiento y tratamiento de información centrados en labores operacionales quedando en un segundo plano la automatización de las tareas de análisis y de gestión, las cuales continuaron en manos de expertos que utilizan en forma restringida las innovaciones tecnológicas, tendencia que se mantiene hasta nuestros días.

El objetivo de este trabajo es superar la brecha tecnológica existente en las organizaciones que usan el paradigma de computación rígida que no ha dado los resultados esperados en los ámbitos de la gestión estratégica, este desafío de automatizar las labores de análisis y gestión requiere sistematizar el conocimiento del negocio, conocimiento que reside en expertos y cuyos paradigmas de trabajo o modelos de negocios están conformados por reglas de proceso que manejan fundamentalmente información imperfecta y cuya interpretación se traduce por sentencias expresadas por medio de lenguaje natural, lo que representa una barrera que entre otras iniciativas encuentra en la teoría de la lógica difusa una excelente manera de superar esta problemática, esto es, aplicando la teoría de los Subconjuntos Difusos que corresponde a una extensión de la teoría clásica de conjuntos, la lógica difusa permite llevar a cabo el proceso de razonamiento en términos imprecisos, a continuación se

revisa en forma somera los principales elementos básicos de esta teoría de subconjuntos difusos y de la lógica difusa.

B2. El Origen de la Lógica Difusa y la Computación Flexible

Los orígenes del término difuso, aplicado a la lógica y a la teoría de conjuntos y sistemas procede de la expresión "fuzzy sets" (conjuntos difusos) acuñada por Lofti A. Zadeh[ZL65], brillante ingeniero eléctrico, nació en 1921 en Baku, una ciudad en el mar Caspio de la antigua república Soviética de Azerbayán. Después de emigrar a Irán, y estudiar en la Universidad de Teherán, llegó a Estados Unidos en donde continuó sus estudios pasando por el M.I.T. la Universidad de Columbia y el Departamento de Ingeniería Eléctrica de la Universidad de California en Berkeley. Por sus numerosas contribuciones a este campo de la ciencia ha recibido multitud de galardones (como la medalla Hamming, 1992) y doctorados honoris causa por multitud de universidades del mundo, entre ellas las de Oviedo (1995) y Granada (1996) en España.

El concepto de Lógica Difusa lo plantea Zadeh en 1965 en su artículo titulado "Fuzzy Sets" [ZL65], complementa su enfoque con sus publicaciones "Outline of a new approach to the analysis of complex systems and decision processes" [ZL73] y "Fuzzy Logic" [ZL88].

Esta teoría y sus conceptos al no contar con capacidad de procesamiento computación suficiente para aplicarla, ve la aparición de sus primeros resultados prácticos nueve años más tarde al presentar en 1974, Ebrahim H. Mandani [ME74], en Inglaterra, la aplicabilidad de la lógica difusa en el campo del control desarrollando el primer sistema de control difuso (Fuzzy Control).

En 1986, Takeshi Yamakawa publica el artículo "Fuzzy controller hardware System", y desarrolla controladores difusos de circuito integrado, y en 1987, se hace evidente el uso masivo de la teoría de conjuntos difusos en oriente, con un hecho relevante, se inaugura en Japón el Ferrocarril Subterráneo Sendai, constituyéndose a esa fecha en el sistema difuso más espectacular puesto en operación a la fecha.

En 1988, con los desarrollos matemáticos del Japonés Michio Sugeno se puede hablar de Ingeniería Difusa.

En 1991, Zadeh introduce el concepto de computación flexible (soft computing), destacando la aparición de las metodologías de computación cuyo objetivo es la tolerancia de la imprecisión y de la incertidumbre incorporando técnicas de la lógica difusa, la neurocomputación, la computación evolutiva, el aprendizaje

automático y el razonamiento probabilística (redes de creencias, teoría del aprendizaje, sistemas caóticos).

Las tesis de Zadeh entroncan con la obra de pensadores de distintas disciplinas que tenían en común una visión semejante de los problemas, alejada de la lógica tradicional. La paradoja del conjunto de Bertrand Russell, el principio de incertidumbre de la física cuántica de Werner Heisenberg, la teoría de los conjuntos vagos de Max Black, otro filósofo cuántico, sin olvidar la fundamental aportación del polaco Jan Lukasiewicz, creador de la lógica multivaluada. Russell y Black utilizaron el término *vagueness* (vaguedad, vago) para referirse a la nueva lógica o a para calificar a los conjuntos en la teorización sobre los mismos, Zadeh prefirió el término "*fuzzy*" (difuso, borroso, difuminado) para denominar a sus conjuntos y a la lógica en la que se apoya su análisis, la diferencia que hace Zadeh del término "vague" y "fuzzy" la ejemplifica Morillas [MR00] como: "Juan regresará en unos pocos minutos" es difuso (impreciso, pero informativo), mientras que "Juan regresará alguna vez" es vago (ambiguo pero no informativo, en el primer caso existe información que puede servir de soporte a una decisión, en el segundo no).

B3. El problema básico

B2.1 La información a partir de datos

Los sistemas de información se muestran robustos para registrar las transacciones o los hechos con los medios disponibles basados en el paradigma de la computación tradicional y es así como los datos se ajustan al modelo registro, campo, valor, modelizado mediante "reglas lógicas clásica" y representan los esfuerzos orientados a capturar la mayor información posible del mundo real mediante modelos semánticos que resuelven las necesidades operacionales y de control básico y no cubre en forma eficiente el problema de tratamiento de información imperfecta característica de los ámbitos del control y la gestión estratégica.

Estos primeros modelos responden al modelo relacional y la información que se puede obtener se clasifica como Información tipo 1, mientras que aquella que depende de múltiples factores y que se puede representar mediante reglas lógicas se denomina Información del tipo 2.

Alternativas al modelo relacional son los modelos semánticos y de orientación al objeto, los desafíos que estos modelos deben superar son la combinación de estructuras de datos y procedimientos interpretativos que utilizados en forma correcta conducen a una conducta inteligente, esto es, derivar información no explícitamente representada.

La tendencia durante años ha sido el acopio de datos recogidos del mundo real que conducen a obtener información del tipo 1, pero el mundo real no siempre nos provee datos suficientes o correctos dando origen a distintos tipos de información imperfecta.

Tipos de Información Imperfecta

Incertidumbre: No existe suficiente información para evaluar la certeza o falsedad de una determinada proposición. Ej. "mañana lloverá".

Imprecisión: No se especifica un valor concreto de un dato sino un conjunto al cual pertenece y depende de la "granularidad" del dominio. Ej. "Juan tiene entre 20 y 30 años".

Vaguedad: el conjunto que se especifica no está bien definido, existiendo vaguedad disyuntiva y conjuntiva.

Vaguedad disyuntiva: el dato real es el único valor del conjunto, Ejemplo: "Juan es joven".

Vaguedad conjuntiva: el dato real tiene como valor todo el conjunto. Ejemplo: "Juan habla idiomas comerciales".

Imprecisión e Incertidumbre: la afirmación es imprecisa y no estamos seguros que sea cierta. Ejemplo: "hay cierta posibilidad que el paciente tenga un "leve" fallo cardiaco".

Tipos de información imperfecta.

Información Errónea, la información es diferente de la existente en el mundo real, el caso más común es la inconsistencia.

Información Incompleta, la información no describe completamente la realidad.

Información Imprecisa, se dispone de un conjunto de valores entre los que se sabe está el verdadero valor del dato, esta información puede ser información nula o información disyuntiva.

Información Incierta, no se tiene seguridad acerca de la certeza de un dato, ya sea por mala calidad de la información o poca fiabilidad de la fuente.

En búsqueda de soluciones que resuelva estas imperfecciones, diferentes autores han trabajado distintas soluciones como los modelos matemáticos de la Teoría de la probabilidad, la Teoría de la

Evidencia, la Teoría de los Factores de Certeza y la Teoría de los Conjuntos Difusos, que revisa Sánchez [44], siendo esta última la de interés para este trabajo.

B3. Los conceptos de la Lógica y de la Lógica Difusa

B3.1. Los conceptos de la Lógica.

La lógica es la ciencia que estudia los principios y métodos para distinguir un razonamiento correcto de otro incorrecto. Se ocupa del conocimiento de la realidad y pertenece a la dimensión práctica de la Filosofía. Leibniz sostiene que la lógica es una especie de lenguaje universal, inteligible para todo el mundo. Por eso, cuando los hombres han de tomar, solos o en grupo, decisiones importantes, es de gran utilidad el uso de la lógica.

La lógica investiga la relación de consecuencia que se da entre una serie de premisas y la conclusión de un argumento correcto. Se dice que un argumento es correcto (válido) si su conclusión se sigue o es consecuencia de sus premisas; de otra forma es incorrecto. En consecuencia, necesitamos de la Lógica para expresarnos en un lenguaje con una sintaxis precisa y una semántica bien definida.

Argumento: es un razonamiento que quiere probar una proposición o afirmación. Debe estar fundamentado, pero sólo será correcto cuando esa fundamentación sea adecuada.

Premisa: es una proposición que se dice con anticipación a algo.

Inducción: es una forma de razonamiento en la que, a partir de observaciones o experiencias determinadas, sacar una conclusión final.

Deducción: es una forma de razonamiento en la que, partiendo de unas premisas y utilizando reglas de derivación (reglas de inferencia) se llega a sacar una conclusión final.

Derivación: es separar cosas de un todo, dividirlo.

Reglas de inferencia: son reglas ya determinadas, por medio de ellas podemos hacer una deducción correcta.

B4. La Lógica difusa

En la Lógica Difusa o Lógica Borrosa que corresponde a la interpretación del término "Fuzzy" de Zadeh, encontramos los siguientes conceptos:

Grado de pertenencia: corresponde a la medida difusa o grado de evidencia o certidumbre sobre la pertenencia de un elemento a un conjunto, el grado de pertenencia se denota con la letra μ que identifica a la función de pertenencia.

Ejemplo:

Universo de edades: $U = \text{edades} = \{5,10,20,30,40,50,60,70,80\}$

Conjuntos borrosos:

bebé= $\{\}$

joven= $\{ 5|1, 10|1, 20|0.8, 30|0.5, 40|0.2, 50|0.1\}$

adulto= $\{ 20|0.8, 30|1, 40|1, 50|1, 60|1, 70|1, 80|1\}$

viejo= $\{20|0.1, 30|0.2, 40|0.4, 50|0.6, 60|0.8, 70|1, 80|1\}$

Altura del conjunto difuso: corresponde al máximo valor de pertenencia obtenido.

Conjunto difuso Normalizado: un conjunto difuso se encuentra normalizado cuando la altura es 1.

Cardinal de un conjunto difuso definido sobre un universo finito: suma de todos los grados de pertenencia:

$$\sum_{x \in U} \mu_A(x) = \text{card}(A)$$

Por ejemplo: $\text{card}(\text{viejo}) = 0.1+0.2+0.4+0.6+0.8+1+1 = 4.1$

Operaciones típicas.

“Un conjunto difuso A es **subconjunto** de otro conjunto difuso B, definido sobre un mismo universo U si el grado de pertenencia de cada elemento del conjunto A es menor o igual a su grado de pertenencia al conjunto B”:

$$A \subset B \text{ si } \mu_A(x) \leq \mu_B(x), \forall x \in U$$

Igualdad y desigualdad:

$$A = B \text{ si } \mu_A(x) = \mu_B(x), \forall x \in U$$

$$A \neq B \text{ si } \exists x \in U, \mu_A(x) \neq \mu_B(x)$$

“**Complemento \hat{A}** de un conjunto difuso A , respecto al universo U , es el conjunto difuso obtenido asignando a cada elemento **del universo** el complemento respecto a la unidad de su grado de pertenencia”:

$$\hat{A} = \{ x \mid \mu_{\hat{A}}(x) = 1 - \mu_A(x), \forall x \in U \}$$

“**Unión** de dos conjuntos difusos, $A \cup B$, como el conjunto difuso obtenido asignando a cada elemento del universo el máximo valor de su grado de pertenencia a cualquiera de los dos conjuntos”:

$$A \cup B = \{ x \mid \mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)), \forall x \in U \}$$

“**Intersección** de dos conjuntos difusos, $A \cap B$, como el conjunto difuso obtenido asignando a cada elemento del universo el mínimo valor de su grado de pertenencia a cualquiera de los dos conjuntos”:

$$A \cap B = \{ x \mid \mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)), \forall x \in U \}$$

Sin embargo, ni contradicción ni tercio excluso, ya que

$$A \cap \hat{A} \neq \emptyset \text{ y } A \cup \hat{A} \neq U,$$

por lo que los conjuntos difusos no forman un álgebra de Boole.

“**Producto cartesiano, $A \times B$** , de dos conjuntos difusos A y B , definidos sobre universos U y V diferentes, como el conjunto difuso obtenido asignando a cada pareja formada por un elemento de U y otro de V , el menor de sus grados de pertenencia al conjunto A y al B respectivamente”:

$$A \times B = \{ (x,y) \mid \mu_{A \times B}(x,y) = \min(\mu_A(x), \mu_B(y)), \forall x \in U, \forall y \in V \}$$

Relaciones difusas

Binaria $R(U,V)$ cuando en la relación sólo intervienen elementos de dos universos:

$$R(U,V) = \{ (x,y) \mid \mu_R(x,y), \forall x \in U, \forall y \in V \}$$

Dominio de una relación borrosa binaria $R(U,V)$ es el conjunto difuso $\text{dom } R(U,V)$ que asigna a cada elemento del conjunto origen U el máximo grado de relación con cualquiera de los elementos del conjunto imagen V .

- **Rango** de una relación borrosa binaria $R(U,V)$ es el conjunto difuso $\text{ran } R(U,V)$ que asigna a cada elemento del conjunto imagen V el máximo grado de relación con cualquiera de los elementos del conjunto origen U .

- Relación binaria compuesta: $R(U,W)=P(U,V) \circ Q(V,W)$ (Composición de relaciones, "**producto matricial max-min**")

- Relación unaria compuesta)

B5. Diferencias entre Lógica difusa y tradicional

1) La lógica tradicional es bivalente, sólo $[0,1]$. La Lógica polivalente según la definición de Lukasiewicz, tiene grados de verdad en el intervalo $[0,1]$, la lógica difusa define varios valores de verdad, varios **grados de pertenencia**.

2) En lógica clásica, **predicados** concretos (mortal, par, mayor...) En lógica difusa, pueden ser concretos o **difusos** (barato, joven, alto, cansado, inteligente, próximo...)

3) la lógica polivalente (bivalente como particular) admite sólo el **cuantificador** universal y el existencial. En difusa, se admiten además **difusos** (muchos, pocos, algunos, la mayoría...)

4) A diferencia de la de predicados de 1er. orden, la lógica difusa admite gran variedad de **modificadores** de predicados (no, muy, mas o menos, casi...). Con éstos se puede hacer que un computador realice cálculos con variables lingüísticas.

5) Apropia para la representación del conocimiento humano. (Para modelar expresiones del lenguaje natural, donde la vaguedad es más deseable que la precisión)

6) Mecanismo de inferencia, expresando las premisas y las conclusiones en forma canónica.

7) Un problema todavía no resuelto con una teoría general en lógica difusa (pero en el que existen propuestas concretas) es la **aproximación lingüística** (cómo encontrar una expresión lingüística cuyo significado sea el del conjunto difuso obtenido de un razonamiento previo.

8) Incluye como caso particular la lógica bivalente, la polivalente, la teoría de probabilidades y la lógica probabilística. (Se puede hablar

del grado de verdad de una sentencia o la probabilidad de ser Verdadero.)

Sentencias borrosas.

Mismo aspecto sintáctico que la lógica de predicados. No se puede construir un sistema axiomático, porque no se puede definir "sentencia válida" (tautología).

Simbología:

- Para las conectivas básicas, los mismos. (Conjunción, disyunción, condicional).

- Para los modificadores (hasta ahora sólo "—"), se usará el adverbio correspondiente (por ejemplo: muy, poco, etc).

- Se usará 'á' para indicar el grado de verdad de una sentencia, con valores entre 0 y 1.

Interpretación de sentencias simples.

- Significado de un **predicado monádico**: conjunto difuso formado por todas las relaciones semánticas (función de pertenencia) del predicado correspondiente con los elementos del universo.

- **Interpretación verdadera de una sentencia simple** (grado de pertenencia del elemento x al conjunto difuso A): $I(A(x)) = \mu_A(x) = \acute{a}$

- Significado de un **predicado n-ario**: relación borrosa formada por todas las relaciones semánticas (función de pertenencia) del predicado correspondiente con las tuplas formadas por los elementos de los distintos universos.

- **Interpretación de sentencias simples con predicados diádicos**:

$$I(R(x,y)) = \mu_R(x,y) = \acute{a}$$

Interpretación de sentencias compuestas

- Regla de disyunción: $I(A(x) \vee B(y)) = \max(\mu_A(x), \mu_B(y))$; $\forall x \in U, \forall y \in V$

- Conjunción dos casos según el papel sintáctico de los predicados:

a) Ambos predicados independientes, o dependientes y la influencia de uno sobre otro es positiva (por ejemplo: joven y saludable):

$$I(A(x) \wedge B(y)) = \min(\mu_A(x), \mu_B(y)); \forall x \in U, \forall y \in V$$

b) Ambos predicados en contradicción (por ejemplo: joven y viejo), coincide con la disyunción.

- Reglas del condicional (según autores):

$$- I(A(x) \rightarrow B(y)) = \max(\min(\mu_A(x), \mu_B(y)), 1 - \mu_A(x)); \forall x \in U, \forall y \in V \text{ (Zadeh)}$$

$$- I(A(x) \rightarrow B(y)) = \min(\mu_A(x), \mu_B(y)); \forall x \in U, \forall y \in V \text{ (Mandani)}$$

$$- I(A(x) \rightarrow B(y)) = \mu_A(x) \cdot \mu_B(y); \forall x \in U, \forall y \in V \text{ (Larsen)}$$

- Interpretación de la sentencia condicional ampliada (si.... entonces.... , si no....):

$$I((A(x) \rightarrow B(y)) \wedge (\neg A(x) \rightarrow C(z))) = \max(I(A(x) \rightarrow B(y)), I(\neg A(x) \rightarrow C(z))) = \max(\min(\mu_A(x), \mu_B(y)), \min(1 - \mu_A(x), \mu_C(z))); \forall x \in U, \forall y \in V, \forall z \in V$$

$$- S(A \vee B) = A \cup B; A \subset U, B \subset U - S(A \wedge B) = A \cap B; A \subset U, B \subset U$$

$$- S(A \wedge B) = A \times B; A \subset U, B \subset V - S(A \rightarrow B) = A \times B; A \subset U, B \subset V$$

$$- S((A \rightarrow B) \wedge (\hat{A} \rightarrow C)) = (A \times B) \cup (\hat{A} \times C); A \subset U, B \subset V, C \subset V$$

Todas estas reglas, formuladas sobre predicados monádicos, son extensibles a predicados diádicos y combinaciones de ambos. Por ejemplo:

$$I(A(x) \rightarrow R(y,z)) = \min(\mu_A(x), \mu_R(y,z)); \forall x \in U, \forall y \in V, \forall z \in W$$

Modificadores lingüísticos. V. Novak propuso:

$$\text{Negación NEG}(\mu(x)) = 1 - \mu(x)$$

Concentración CON($\mu(x)$) = $\mu^2(x)$ (Reduce el grado de pertenencia: "muy")

Dilatación DIL($\mu(x)$) = $2\mu(x) - \mu^2(x)$ (al revés (lo amplía): "más o menos")

Intensificación INT($\mu(x)$) = $2\mu^2(x)$ si $0 \leq \mu(x) \leq 0.5$; $1 - 2(1 - \mu(x))^2$ si $\mu(x) > 0.5$ (se combina con alguna otra operación, p.e. con Concentración sería "bastante").

- Otros autores coinciden con Novak en NEG y CON, pero: $DIL(\mu(x)) = \sqrt{\mu(x)}$ Y de esta forma se verifica que $CON(DIL(\mu(x))) = DIL(CON(\mu(x))) = \mu(x)$

- Reglas de modificación para interpretación de sentencias simples con los adverbios más comunes:

“no”: $I(\neg A(x)) = NEG(\mu A(x)) = 1 - \mu A(x)$

“muy”: $I(muy A(x)) = CON(\mu A(x)) = \mu^2(x)$

“hiper”: $I(hiper A(x)) = \mu A(x)$ $CON(\mu A(x)) = \mu^3(x)$

“más o menos” y “algo” $I(más o menos A(x)) = I(algo A(x)) = DIL(\mu(x)) = 2\mu(x) - \mu^2(x)$

“Aproxim.” y “casi” $I(aproximadamente A(x)) = I(casi A(x)) = DIL(DIL(\mu(x))) =$

“Bastante” $I(bastante A(x)) = INT(CON(\mu A(x)))$

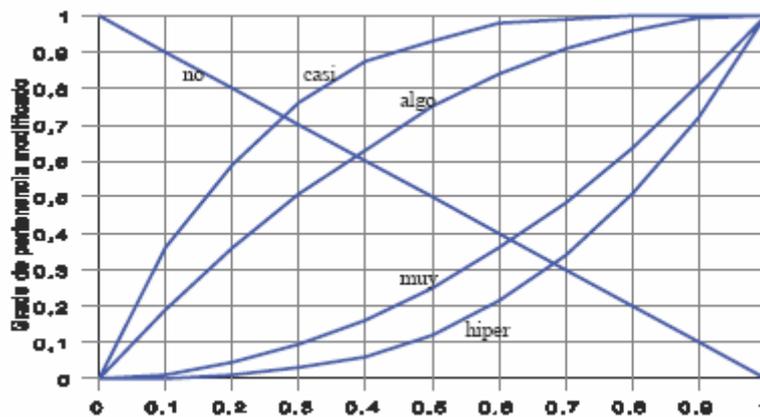


Fig. B.1. Grado de Pertenencia

Razonamiento categórico: las premisas no contienen cuantificadores difusos (por ejemplo: la regla del Modus Ponens Generalizado).

Razonamiento disposicional: una o más premisas pueden contener implícita ó explícitamente el cuantificador difuso “normalmente”.

Razonamiento silogístico: las premisas contienen cualquier cuantificador difuso.

Razonamiento cualitativo: el problema se modela con un conjunto de relaciones borrosas del tipo “Si condiciones ENTONCES acciones”, empleado en aplicaciones de control.

Seis Reglas básicas de inferencia en razonamiento categórico, que permiten obtener conclusiones a partir de un conjunto de premisas (aunque estas conclusiones no son precisas sino aproximadas), y que constituyen por tanto la base del **Razonamiento Aproximado:**

1) **Principio de herencia:** "si el conjunto difuso A, que está asociado a un predicado, es un subconjunto del conjunto difuso B, cualquier elemento del universo que tenga la propiedad presentada por el conjunto A también hereda la propiedad presentada por el conjunto B"

$$\begin{aligned} &A(x) \\ &A \subset B \\ &B(x) \quad \forall x \in U, A \subset U, B \subset U \end{aligned}$$

2) **Regla de la intersección:** "si un elemento del universo tiene las propiedades representadas por los conjuntos difusos A y B (definidos sobre el mismo universo), también tiene la propiedad representada por la intersección de ambos conjuntos"

$$\begin{aligned} &A(x) \\ &B(x) \\ &(A \cap B)(x) \quad \forall x \in U, A \subset U, B \subset U \end{aligned}$$

3) **Regla del producto cartesiano:** "si dos elementos de distintos universos tienen las propiedades representadas por los conjuntos difusos A y B, la pareja formada por ellos tiene el grado de relación que le corresponde en el producto cartesiano de ambos conjuntos"

$$\begin{aligned} &A(x) \\ &B(y) \\ &(A \times B)(x,y) \quad \forall x \in U, A \subset U, \forall y \in V, B \subset V \\ &[\text{siendo } \mu_{A \times B}(x,y) = \min(\mu_A(x), \mu_B(y))] \end{aligned}$$

4) **Regla de la proyección:** "de una pareja de elementos que cumpla una relación borrosa se puede concluir que el primer elemento tiene la propiedad representada por el dominio de la relación"

$$\begin{aligned} &R(x,y) \\ &\text{dom } R(x) \quad \forall x \in U, \forall y \in V, R \subset U \times V \end{aligned}$$

5) **Regla de la composición:** "de un elemento que cumpla una propiedad, representada por un conjunto difuso A, y que esté relacionado con otro elemento de un universo diferente se puede concluir que éste segundo elemento tiene la propiedad representada por la composición de la relación unaria A y de la relación binaria R"

$$\begin{aligned} &A(x) \\ &R(x,y) \\ &(A \circ R)(y) \quad \forall x \in U, \forall y \in V, A \subset U, R \subset U \times V \end{aligned}$$

6) **Modus Ponendo Ponens Generalizado (MPG):**

$$\begin{aligned}
&A(x) \\
&B(x) \rightarrow C(y) \\
&(A \text{ o } (B \times C))(y) \quad \forall x \in U, \forall y \in V, A \subset U, B \subset U, C \subset V
\end{aligned}$$

- Una característica importante del MPG que no posee el MP de la lógica bivalente, es que el antecedente $B(x)$ no tiene por qué coincidir con la premisa $A(x)$.

- Se puede considerar como un caso particular de la regla de composición, siendo $R = B \times C$ la relación borrosa generada por el condicional en el universo $U \times V$.

A10. El concepto de subconjunto difuso

Como es sabido, de la teoría de conjuntos clásicos, un conjunto es una colección de objetos o elementos bien especificados que poseen una propiedad común, la cual puede ser definida por enumeración de sus elementos, por una descripción analítica de una propiedad que lo caracteriza, o usando una función característica también llamada función de pertenencia. Si analizamos el paradigma de la forma clásica de organizar los datos y manipularlos, vemos que su característica responde al modelo clásico de la lógica bivaluada, la que se puede expresar como una función de pertenencia:

$$\mu.M(x) = \begin{cases} 1 & \text{si } x \in M \\ 0 & \text{si } x \notin M \end{cases}$$

Es decir, una función escalón centrada en el valor umbral de decisión.

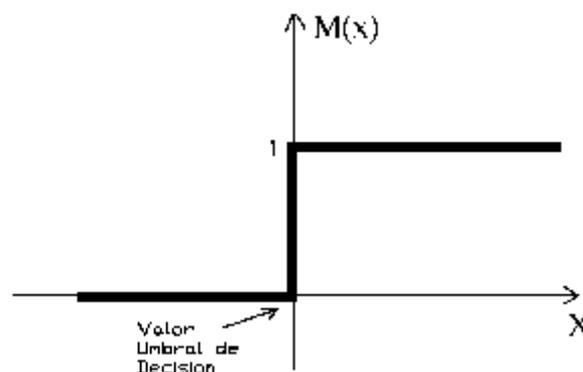


Fig. B.2. Función Escalon

Este modelo clásico de organizar y tratar los datos no resuelve las imperfecciones señaladas en los apartados anteriores, restricciones que sólo permite extraer de los datos información no conocimiento,

esta brecha existente entre la forma de organizar y manipular datos y la capacidad de aprender y generar conocimiento que emplean los seres humanos, es la brecha que pretende cubrir los esfuerzos que se realizan en el ámbito de la lógica del razonamiento aproximado o de los conjuntos y la lógica difusa.

Como ya se ha planteado, la teoría de los subconjuntos difusos como una extensión de la teoría clásica de conjuntos en que la función característica que define la pertenencia de un elemento del universo de referencia al conjunto, tiene como conjunto al intervalo real $[0,1]$, en lugar del conjunto discreto $\{0,1\}$.

Mediante notación matemática se define en [8], un Conjunto Difuso B como:

$$B = \{ (x, \mu_B(x)) / x \in X \}$$

$$\mu_B: X \rightarrow [0,1]$$

La representación de la función de pertenencia de un elemento a un Conjunto Difuso se representa según la figura siguiente:

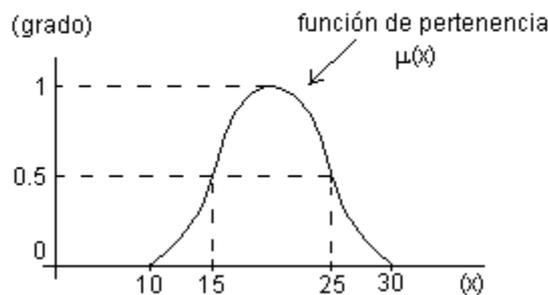


Fig. B.3. Ejemplo de función de pertenencia a un Conjunto Difuso.

Alfa – Corte

Un concepto de mucha utilidad en la teoría de subconjuntos y lógica difusa es el conjunto de α -corte, o grado de presunción, nivel o umbral α , que se define algebraicamente como:

$$A_\alpha = \{x \in U : \mu_A(x) \geq \alpha\}, \alpha \in [0,1]$$

Que significa que es un conjunto que contiene todos los valores de x con un valor de compatibilidad, pertenencia, presunción o certeza de al menos α , si sólo se consideran los valores de x mayores que α se denomina α -corte estricto, fuerte, si $\alpha = 1$ al conjunto $A_\alpha = 1$ se suele llamar Kernel de A o Núcleo de A.

De esta forma α -corte es un Intervalo real para el cual todos sus elementos pertenecen al número difuso con un nivel de confianza mayor o igual a α .

Algebraicamente:

$$\chi \in R / A(x) \geq \alpha$$

Soporte:

Alfa-corte de nivel cero. Representado por: $A(0)$

Moda o uno-corte:

Alfa-corte de nivel uno. Representado por: $A(1)$

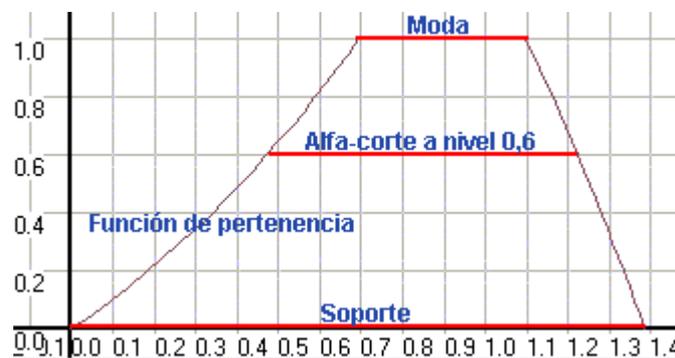


Fig. B.4. Alfa Corte

Operaciones lógicas sobre conjuntos difusos

Al igual que en la lógica tradicional, los operadores lógicos que utiliza la Lógica Difusa (AND, OR, etc.) se definen también usando tablas de verdad, pero mediante un "principio de extensión" por el cual gran parte del aparato matemático clásico existente puede ser adaptado a la manipulación de los Conjuntos Difusos y, por tanto, a la de las variables lingüísticas.

La operación más importante para el desarrollo y creación de Reglas Lógicas es la implicación, simbolizada por " \rightarrow " que representa el "Entonces" de las reglas heurísticas: Si (...) Entonces (\rightarrow) (...).

Así, en la Lógica Difusa hay muchas maneras de definir la implicación. Se puede elegir una "función (matemática) de implicación" distinta en cada caso para representar a la implicación.

La última característica de los sistemas lógicos es el procedimiento de razonamiento, que permite inferir resultados lógicos a partir de una serie de antecedentes. Generalmente, el razonamiento lógico se basa en silogismos, en los que los antecedentes son por un lado las proposiciones condicionales y las observaciones presentes por otro.

Los esquemas de razonamiento utilizados son "esquemas de razonamiento aproximado", que intentan reproducir los esquemas mentales del cerebro humano en el proceso de razonamiento. Estos esquemas consistirán en una generalización de los esquemas básicos de inferencia en Lógica Binaria (silogismo clásico).

Tan importante será la selección de un esquema de razonamiento como su representación material, ya que el objetivo final es poder desarrollar un procedimiento analítico concreto para el diseño de controladores difusos y la toma de decisiones en general.

Las operaciones lógicas que se pueden establecer entre conjuntos difusos son la intersección, la unión y el complemento, igual que las que usamos en lógica bivaluada. Mientras que el resultado de operar dos conjuntos 'abruptos' es un nuevo conjunto 'abrupto', las mismas operaciones con conjuntos difusos nos darán como resultado otros conjuntos también difusos. Dado que la lógica difusa es una extensión de la bivaluada, las nuevas operaciones que explicaremos para intersectar o unir conjuntos difusos son también aplicables a la lógica bivaluada obteniendo idénticos resultados.

En lógica difusa hay muchas maneras de definir estas operaciones. Cualquier operación que cumpla las restricciones de una **T-Norma** puede ser usada para intersectar, igual que cualquier **S-Norma** o **T-Conorma** puede ser usada para unir conjuntos difusos. Las T-Normas especifican un conjunto de condiciones que deben reunir aquellas operaciones que deseen ser usadas para intersectar conjuntos, mientras que las T-Conormas o S-Normas hacen lo propio para las uniones.

Norma Triangular (t-norma)

Sean $c, a, b, e, d \in [0,1]$ La intersección se generaliza por medio de la t-norma, que se escribe $a * b$ y se define por las siguientes propiedades:

$$T1 \quad a * b = b * a \quad (\text{Conmutatividad})$$

$$T2 \quad (a * b) * c = a * (b * c) \quad (\text{Asociatividad})$$

$$T3 \quad \text{si } a \leq c \text{ y } b \leq d \Rightarrow a * b \leq c * d \quad (\text{Monotonicidad})$$

$$T4 \quad \forall a \in [0; 1] \exists 1 \text{ t.q. } a * 1 = a$$

Es fácil comprobar que $a * b = \min(a; b)$ cumple las 4 propiedades de la t-norma.

En muchas aplicaciones se busca establecer la relación *Causa-Efecto*, y por tanto debe emplear un tipo de implicación que sea verdadera sólo si tanto antecedente como consecuente son verdaderos (**AND : T-Normas**)

Norma Triangular Complementaria (T-Conorma o S-norma)

La unión se generaliza por medio de la s-norma $a _ b$ definida por las propiedades:

$$T1 \quad a _ b = b _ a \quad (\text{Conmutatividad})$$

$$T2 \quad (a _ b) _ c = a _ (b _ c) \quad (\text{Asociatividad})$$

$$T3 \quad \text{si } a \leq c \text{ y } b \leq d \Rightarrow a _ b \leq c _ d \quad (\text{Monotonicidad})$$

$$T4 \quad \forall a \in [0; 1] \exists 1 \text{ t.q. } a _ 0 = a$$

Estos es, el 1 es un elemento idéntico para la t-norma y el cero es para s-norma.

Las operaciones lógicas más usuales son el **complemento**, la **unión**, la **intersección**, de las cuales se ilustran a continuación con ejemplos publicados por López [LJ00], Duarte[DO97] y Pérez[PG02].

B6. Definición de conceptos relacionados con la extracción de conocimiento

Cardinal Difuso: como es sabido, en la teoría de conjuntos clásicos, el cardinal de un conjunto representa el número de elementos que contiene ese conjunto, en el caso del Cardinal Difuso, este corresponde a un conjunto difuso sobre enteros no negativos.

Cardinal Relativo: el Cardinal relativo de dos conjuntos $X1$ y $X2$, se define como el porcentaje de elementos de $X2$ que pertenecen a $X1$.

Entropía de un Conjunto Difuso: la entropía de un conjunto difuso es una medida del grado de imprecisión, es decir, como una medida de lo difuso que es el conjunto difuso, se puede interpretar como una medida de la cantidad de información que se necesita para convertir un conjunto difuso en uno clásico.

Equipotencia: se dice que dos conjuntos son equipotentes cuando ambos poseen el mismo número de elementos, por tanto, el mismo cardinal.

Sentencia Cuantificada: son sentencias de lenguaje natural en las que se utilizan etiquetas para realizar afirmaciones acerca de la cantidad o porcentaje de elementos de un conjunto que verifican una cierta propiedad, para Zadeh este tipo de sentencias es considerada importantes por la frecuencia de uso humano, y para el existen dos tipos de sentencias cuantificadas:

Sentencias de tipo I: Q de los X son A

Sentencias del tipo II: Q de los D son A,

Siendo Q la etiqueta lingüística, X un conjunto finito, y A y D son dos subconjuntos difusos de X que representan una propiedad imprecisa., un ejemplo del sentencia cuantificada tipo I es "Alrededor de 30 estudiantes son

jóvenes” y un ejemplo del tipo II es: “La mayoría de los estudiantes eficientes son jóvenes”.

De esta manera las sentencias cuantificadas pueden ser de dos formas:

$Qx \text{ `s son } B \rightarrow$ Ejemplo: “la mayoría de los alumnos son jóvenes”, el grado de verdad de este tipo de sentencias se evalúa como:

$Q(r)$ donde $r = \text{Card}(B) / \text{Card}(X) = (\sum_i B(x_i)) / n$, siendo X el universo del discurso de n elementos.

$Q A x \text{ `s son } B \rightarrow$ Ejemplo: “la mayoría de los alumnos Altos son jóvenes”, el grado de verdad de este tipo de sentencias se evalúa como:

$Q(r)$ donde $r = \text{Card}(A \cap B) / \text{Card}(A)$.

Cuantificador Lingüístico: es un caso particular de una etiqueta lingüística que representa cantidades o porcentajes impreciso.

Cuantificadores Lingüísticos o Difusos:

Se usan para medir (o cuantificar) la cantidad o la proporción de objetos o elementos que cumplen o satisfacen cierta condición.

En lógica clásica existen dos muy importantes:

\forall (todo): se refiere a todos los elementos u objetos

\exists (existe): se refiere al menos a uno de los elementos u objetos.

Usando lógica difusa existen más clasificados en dos categorías:

Cuantificadores Absolutos: se refiere a una única cantidad determinada para medir si esa cantidad son “muchos”; “pocos”; “muchísimos”; “aproximadamente entre 5 y 7” “aproximadamente más de 34”; “aproximadamente 9”..., para evaluar la verdad del cuantificador se requiere de un cuantificador absoluto, se necesita una única cantidad, estos Cuantificadores absolutos, representan cantidades o intervalos de cantidades difusas y que por su semántica se representan por medio de conjuntos difusos sobre números enteros no negativos, por ejemplo: alrededor de 2, aproximadamente entre 1 y 3.

Cuantificadores Relativos: se refiere a una proporción de elementos respecto del total de los que existen, por ejemplo: “La mayoría”; “La minoría”; “casi todos”; “casi ninguno”; “Aproximadamente la mitad”..., para evaluar la verdad se requieren

dos cantidades: la cantidad de elementos que cumplen la condición y el número total de elementos existentes, estos **Cuantificadores relativos** representan porcentajes difusos del total, su representación semántica se define mediante un conjunto difuso sobre un intervalo real $[0,1]$, por comodidad, aunque en la práctica se utilizan tan solo valores racionales, por ejemplo: “al menos la mitad”, “la mayoría”, “todos”, “existe”, “aproximadamente la mitad”. Hay dos formas de interpretar la semántica del cuantificador relativo, con un sentido exclusivo y con un sentido amplio, por ejemplo si es exclusivo, el cuantificador “la mayoría” no incluiría a “todos”, en el caso amplio si lo considera.

Según Zadeh, los **cuantificadores difusos** se representan como conjuntos difusos con dominio subyacente en los números reales.

El **dominio subyacente** esta limitado de limitado dependiendo del tipo de cuantificador:

Cuantificadores Difusos Absolutos: $Q_{abs}: R^* \rightarrow [0,1]$

Cuantificadores Difusos Relativos: $Q_{rel}: [0,1] \rightarrow [0,1]$, en los relativos, el cuantificador se aplica a la división del número de elementos que cumplen la condición entre el número de elementos totales.

Granularidad: es una propiedad relacionada con el número de etiquetas consideradas en el conjunto, una granularidad baja o gruesa (Coarse), se corresponde con un número bajo de etiquetas, el dominio esta poco particionado y tiene el inconveniente que puede perder expresividad, por el contrario una granularidad alta o delgada (fine) se asocia con un número alto de etiquetas, una granularidad demasiado alta puede provocar un aumento de la complejidad en la descripción del dominio.

Normalmente se usa un conjunto pequeño de valores para una variable lingüística, esto define su **Granularidad:**

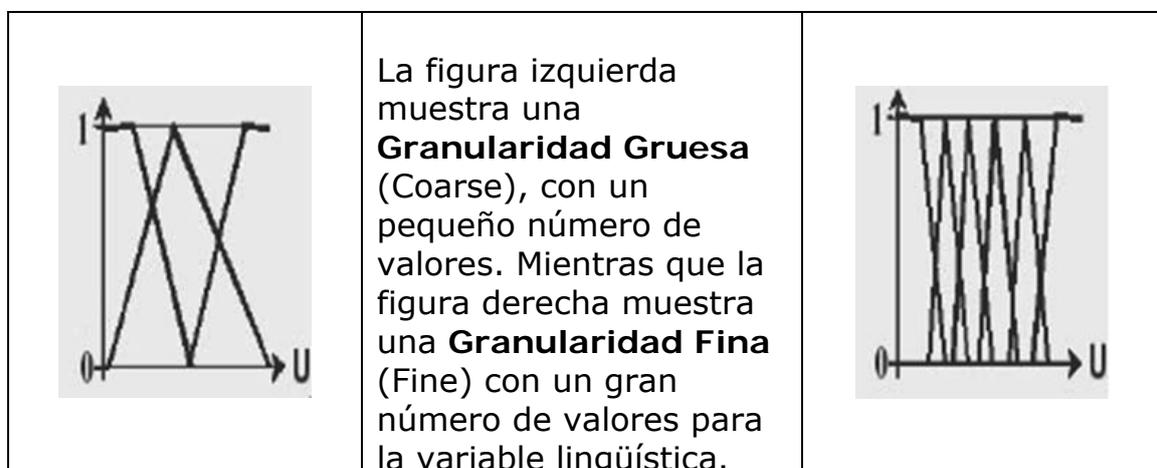


Fig. B.4. Granularidad

Anexo C Variables y Etiquetas Lingüísticas.

C1. Variables Lingüísticas.

Una variable lingüística (linguistic Variable), es una variable cuyos valores son palabras o sentencias (no números). Con frecuencia se desea describir el estado de un objeto o fenómeno, para ello se utiliza una variable cuyo valor hace la descripción, por ejemplo la temperatura, el nivel de comodidad, el tamaño. Una variable lingüística admite que sus valores sean etiquetas lingüísticas, que son términos lingüísticos definidos como conjuntos difusos (sobre cierto dominio subyacente).

Por ejemplo la temperatura “Calida” o “Aproximadamente 25°C”, el dominio subyacente, es un dominio numérico: Los grados centígrados.

Un valor concreto, crisp (25°C, por ejemplo), es en general más específico que una etiqueta lingüística. Es un punto del conjunto, mientras que una etiqueta lingüística es una colección de puntos (temperaturas posibles).

Hay variables cuya defición es más compleja porque se mueven en dominios subyacentes poco claros y no es natural trasladarlos a valores numéricos, por ejemplo: limpieza, calidad.

Las variables lingüísticas son elementos fundamentales en cualquier sistema de lógica difusa. En ellas se combinan categorías subjetivas que describen el mismo concepto, así, para el caso de la variable altura existirán las categorías: *bajo, mediano, alto y muy alto*, que son llamados términos lingüísticos y representan los posibles valores de una variable lingüística.

Variable Lingüística

La definición de Zadeh considera un conjunto de cinco elementos para definir una variable lingüística, a saber: $\langle N, U, T(N), G, M \rangle$, donde:

N es el nombre de la variable

U es el dominio subyacente

T(N) es el conjunto en términos de etiquetas que puede tomar N

G es una gramática para generar las etiquetas de T(N): por ejemplo: “bajo”; “Normal”; “Extremadamente normal”, “no muy alto”; “alto”; “muy alto”...

M es una regla semántica que asocia cada elemento T(N) con un conjunto difuso de U de entre todos los posibles: $M: T(N) \rightarrow F(N)$

Gramática G

Normalmente los símbolos terminales incluyen:

- Un conjunto de **términos primarios** (primary term): Bajo; Alto...
- Un conjunto de **modificadores** (hedges): Muy; Más o Menos; completamente; especialmente; Cerca de; Aproximadamente...
- Un conjunto de **conectivos lógicos**: normalmente NOT; AND y OR, estos conectivos también reciben el nombre de modificadores (Modifiers).

Grado de Pertenencia:

Es el grado con el cual una entrada bien definida es compatible con una función de pertenencia, puede tomar valores entre 0 y 1. Por ejemplo, el grado de pertenencia de x al conjunto difuso alto (A) es representado por la función $A(x)$, donde x es un valor numérico de altura dentro del universo U ($x \in U$). El rango de A es cualquier valor entre 0 y 1, según represente algún valor entre ningún o total grado de pertenencia al conjunto difuso.

Termino

Es una categoría subjetiva de una variable lingüística y consecuentemente, es el nombre descriptivo usado para identificar una función de pertenencia. Tal como las variables algebraicas toman valores numéricos, las variables lingüísticas toman como valores términos lingüísticos

Entradas bien definidas (entradas crisp)

Son los diferentes valores discretos de la variable del sistema por ejemplo las alturas de un grupo de personas 1.60m; 1.75m; 1.80m, etc. En oposición al concepto difuso, lo crisp, definido, nítido o preciso no representa ninguna incerteza o imprecisión.

Rango o Dominio

Es el intervalo sobre el cual se define una función de Pertenencia. Por ejemplo, una función de pertenencia Alto podría tener un dominio de 1.60m a 1.90m y su rango sería 0.30m, la siguiente figura, muestra gráficamente los conceptos que se acaban de revisar.

C2. Etiquetas Lingüísticas.

De acuerdo a lo indicado por Serrano en [54], en ocasiones resulta necesario describir una propiedad o el estado de un objeto o fenómeno mediante una expresión en lenguaje natural. Para ello se ha de usar lo que se conoce como una variable lingüística, la cual admite el que sus valores sean etiquetas lingüísticas.

Una etiqueta lingüística es un término extraído del lenguaje natural mediante el que expresamos una valoración imprecisa que, en ciertos casos, puede tener una representación mediante un conjunto difuso. Por ejemplo, “*profundo*” es una etiqueta lingüística, mientras que el intervalo $[15,25]$ (que puede interpretarse como un conjunto difuso) no lo sería. Este concepto fue desarrollado por Zadeh y es muy utilizado cuando el universo de discurso es algún subconjunto de la recta real

Una de las posibles aplicaciones del concepto de etiqueta lingüística y de la que se hará uso más adelante es la definición de conjuntos de etiquetas sobre dominios. La siguiente figura nos muestra un ejemplo, en el que se ha definido un conjunto de etiquetas $\{Muy\ bajo, Bajo, Medio, Alto, Muy\ alto\}$, sobre un dominio numérico que en el presente caso podría corresponder a una característica tal como la Altura de un individuo, expresada en metros.

A la representación de una etiqueta lingüística como conjunto difuso se le denomina Representación semántica.

Existen conjuntos de variables cuya definición es más compleja porque se mueven en dominios subyacentes poco claros, que no resultan fáciles de trasladar a valores numéricos: Sabor de una comida, Color del pelo, etc.

La utilidad que una etiqueta lingüística pueda incluir muchos valores posibles, según Zadeh, ayuda a caracterizar fenómenos que son complejos de definir. Es un medio de trasladar conceptos o descripciones lingüísticas a descripciones numéricas que pueden ser tratadas automáticamente, relaciona o traduce el proceso simbólico a proceso numérico. Usando el principio de extensión, muchas herramientas ya existentes, pueden ser extendidas para manejar variables lingüísticas, obteniendo las ventajas de la lógica difusa en gran cantidad de aplicaciones.

D. La Industria de las Grandes Tiendas

A continuación se presenta un análisis de la industria de las grandes tiendas o grandes empresas dedicadas a la comercialización de vestuario y hogar en el formato de tiendas por departamentos. Para este apartado se ha considerado pertinente utilizar y adaptar el trabajo realizado por Cuesta en [CP06], por su estructura y completitud. Para esta adaptación se han tenido en cuenta los siguientes dos aspectos: Primero Chile es un país emisor de inversiones a diferencia del trabajo de Cuesta que muestra la perspectiva de España como receptor de inversiones y en segundo lugar el ámbito de este trabajo son las grandes tiendas de venta por departamentos en lugar del ámbito de Cuesta que se refiere a los supermercados.

D1. El perfil de las grandes tiendas en Chile

La industria de las grandes tiendas en Chile según las recientes publicaciones de organismos especializados como por ejemplo lo señalado por Jarufe en [JR05], es liderada cuatro empresas que en 2005 alcanzaron ventas por USD\$ 3.194 millones y cuyas participaciones son las siguientes:

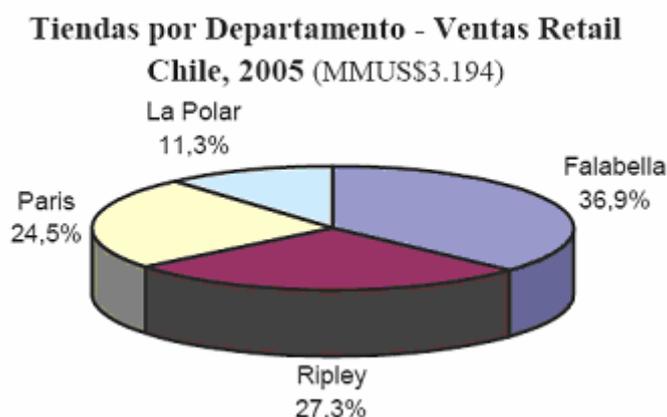


Fig. D1 Tiendas x departamento – Ventas 2005, fuente [JR05]

El perfil de cada una de estas empresas es el siguiente:

D1.1. Falabella:

Falabella nace en 1889, como una sastrería en el centro de Santiago (Ahumada). En 1937 esta empresa familiar se convierte en sociedad anónima y con una nueva estructura comienza formarse una tienda completa para caballeros, mientras que a partir de 1952, se hace lo mismo con la moda femenina. Con el correr de los años, se agregan locales destinados al vestuario de jóvenes, niños, todos ubicados en el centro de Santiago. Durante la década del 70, comienza el desarrollo hacia regiones, partiendo en Concepción, para luego incursionar en otras ciudades del país y ampliar su rubro a las líneas de electrohogar (decoración y electrónica).

En el año 1980 se crea la Promotora CMR, como un sistema de crédito a los clientes. Durante el año 1980, Falabella se asocia con otras sociedades para incursionar en el área inmobiliaria, participando en la construcción de centros comerciales en el país. En 1993, la empresa inicia su expansión internacional hacia Argentina abriendo su primera tienda en Buenos Aires en 1999. En 1995, comienza a operar en Perú bajo la marca Saga Falabella, mientras que en 2002 comienza a incursionar en el formato hipermercados a través de Tottus.

En 1996, la empresa se abrió a la bolsa de valores. En 1997 Falabella (33%) se asocia con la norteamericana The Home Depot (66%) para instalar en Chile una cadena de tiendas de mejoramiento del hogar, alcanzando la totalidad de dicho negocio a fines de 2001. En octubre de 2003 Falabella se fusiona con Sodimac (grupo Dersa de la familia del Río), empresa líder en el rubro de distribución de materiales de construcción y mejoramiento del hogar. Este hecho lleva a que los activos de la compañía crezcan en más de un 20%, mientras que las ventas consolidadas aumentan en cerca de un 50%. Así, la empresa quedó controlada por las familias Solari y Del Río, las que mantienen un pacto de accionistas en un directorio compuesto por 5 miembros de la familia Solari, 2 representantes de Dersa y 2 independientes.

En 1998 compra la licencia ING Bank (Chile) creándose el Banco Falabella. En 1999 Falabella adquiere el 20% de Farmacias Ahumada y además se inicia en el negocio de comercio por internet.

Finalmente en el año 2004, Falabella ingresa al rubro de alimentos con la adquisición del 88% de supermercados San Francisco y en el año 2005, la empresa aumenta su participación en Sodimac Colombia de 35% a 49% y lanza al mercado la tarjeta CMR.

La propiedad de Falabella se concentra en las familias Solari (65%) y Del Río (22,3%), mientras que inversionistas alcanzan un 12,7% del total accionario.

De acuerdo a los ingresos consolidados de Falabella en 2005, un 81% proviene de operaciones en Chile, desglosado en un 91% en negocio retail y 9% en negocio financiero, un 14% en Perú (considerando los ingresos de retail y financiero), Argentina un 4%(incluyendo ingreso retail y financiero) y Otros el porcentaje restante.

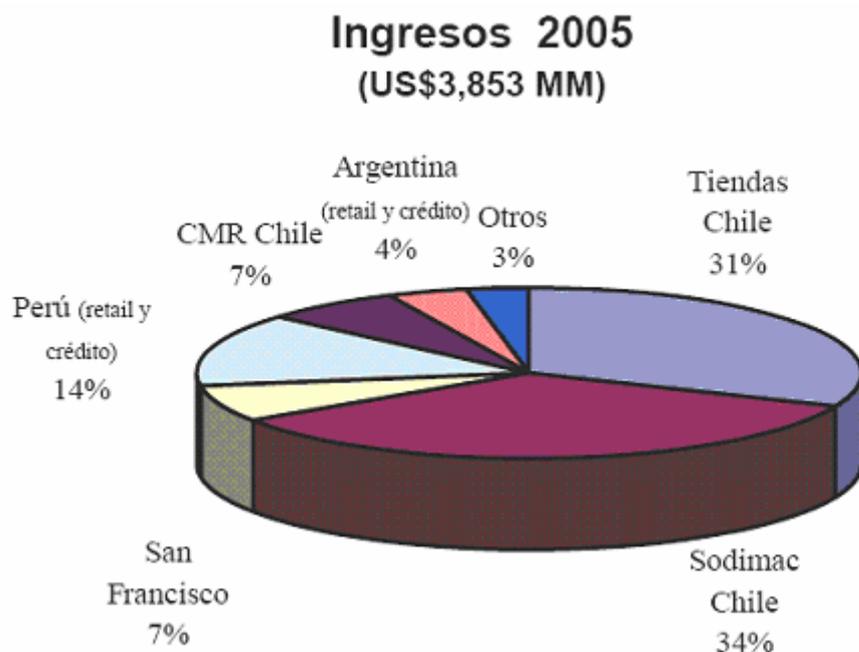


Fig. D.2. Ingresos de Falabella en 2005, fuente: Memoria Anual Falabella

D1.2. Ripley:

Ripley es hoy una de las mayores compañías del sector retail en Chile y Perú. El principal negocio es la venta al detalle de vestuario, accesorios y productos para el hogar a través de los distintos formatos de tiendas por departamento. Al mismo tiempo participa en el negocio financiero a través del crédito directo a sus clientes por medio de la Tarjeta Ripley y del Banco Ripley. Asimismo, Ripley participa en la propiedad de centros comerciales y en otros negocios asociados al retail como viajes y corretaje de seguros.

En definitiva, Ripley es un modelo de retail centrado en el cliente y enfocado a entregar las mejores soluciones a las necesidades de los consumidores.

Reseña Histórica:

- 1956 Calderón Confecciones, negocio principalmente de confección y venta por tienda al por menor de ropa masculina, inaugura el primer establecimiento.
- 1964 En pleno centro de Santiago, se inaugura la primera tienda Ripley.
- 1976 Se comienza a operar con un sistema de crédito semiautomático en cada sucursal.
- 1978 Surge la razón social Comercial Eccsa S.A (Establecimientos Calderón Confecciones S.A.), que agrupaba a todas las tiendas existentes en esa época y que se mantiene hasta hoy. De esta forma se separó la actividad comercial de la productiva, que seguía a cargo de Calderón Confecciones.
- 1993 Se inaugura la primera tienda en un mall, Ripley Parque Arauco, lo cual implicó un profundo y exitoso cambio de imagen y reposicionamiento. Además, nace Ripley Novios, como nuevo servicio.
- 1997 Se inaugura la primera tienda fuera de Chile, Ripley Jockey Plaza, en Lima Perú. Además se crea la empresa CAR S.A., la cual nace para administrar el negocio del crédito, y fortalecer los servicios al cliente.
- 1999 Se autoriza la operación de Financiera Cordillera (Financor) en Perú, la que fue creada para ofrecer tarjetas de crédito a los clientes de Ripley.
- 2000 Ripley se incorpora al e-business, por medio del sitio www.ripley.cl y se autoriza la operación de la Corredora de Seguros Ripley. En Perú se da inicio a las operaciones de Financor.
- 2002 Banco Ripley comienza sus operaciones. Se forma la alianza Lan-Ripley en el negocio de viajes
- 2003 Ripley Chile S.A. realiza la primera emisión de bonos en el mercado local.
- 2005 El 14 de julio Ripley Corp S.A. realiza su exitosa apertura bursátil de un 15% de la propiedad final en la Bolsa de Comercio de Santiago. Se inauguran dos nuevas sucursales chilenas, la tienda de Ripley Crillón, y Portal Temuco; además se inauguró, durante diciembre, la tienda de Chorrillos en Perú. Banco Ripley durante el año concretó la apertura de 6 nuevas sucursales, además de la creación de un Centro de Inversiones en Santiago.

Tal como se menciona, el año 1993, fue el año clave en la historia de Ripley, la Compañía entró al Mall Parque Arauco y comenzó un cambio de imagen y un fuerte reposicionamiento de marca. Durante los últimos años el negocio de las tiendas por departamentos de Ripley ha tenido un gran crecimiento.

La exitosa estrategia de Ripley considera una fuerte interacción entre esta área de negocios y otras que complementan y potencian su actividad, como es el caso del negocio financiero a través de la Tarjeta Ripley.

Ripley tiene 31 tiendas en Chile, 15 en Santiago y 16 en regiones con una superficie total de 227.909 mts² construidos.

Ripley cuenta en sus tiendas con una serie de servicios adicionales, que generan un grado de fidelización adicional con sus clientes.

Las listas de Regalo actualmente vigentes, tales como Lista de Novios, Lista del Bebé, Lista de Decoración y Listas de Viajes de Estudio (estas últimas dos, creadas el año 2005) persiguen focalizar las compras de sus clientes para acontecimientos muy importantes en sus vidas. Para complementar estas listas, en el 2005 se creó un servicio de decoración que permite brindar una asesoría integral, con especialistas en la materia para el cambio y ornamentación de la casa.

Ripley también ha creado canales de venta no tradicionales que complementan su oferta, tal es el caso del canal de Venta Virtual (ripley.cl) que genera diariamente un comercio electrónico a distancia. Pero no solo en su crecimiento en ventas y aumento de clientes este canal se ha desarrollado en los últimos años, sino que también en su capacidad y alcance para comunicarse con los clientes. Todo lo cual ha permitido potenciar y difundir todo el giro de negocio que Ripley ofrece (seguros, viajes, Tarjeta Ripley, Novios, etc), y también generar un excelente medio de publicidad de sus campañas comerciales para los miles de clientes que diariamente entran a navegar por sus páginas.

Asimismo, la venta corporativa ha crecido enormemente en los últimos años. Este canal se preocupa de satisfacer las necesidades de las empresas fundamentalmente radicadas en beneficios a sus funcionarios, en promociones a sus clientes o en las necesidades de su propio abastecimiento.

INGRESOS DE RIPLEY EN 2005

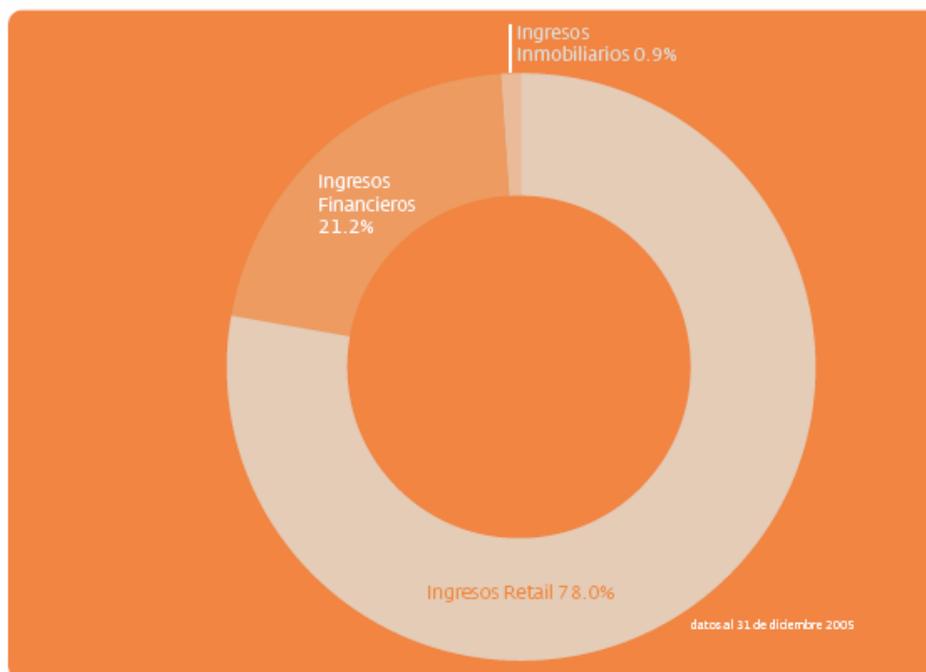


Fig. D.3. Distribución de Ingresos de Ripley en 2005, Fuente: Memoria Anual Ripley

D1.3. Almacenes Paris:

Almacenes Paris, empresa ligada a la familia Gálmez, se fundó en 1900 con el nombre de "Muebles Paris", dedicada en ese entonces sólo al rubro de fabricación y comercialización de muebles. La primera tienda se ubicó en calle Alameda y a comienzos de la década de los 50, la Mueblería Paris incluyó vestuario y otros relacionados, cambiando al mismo tiempo su nombre a Almacenes Paris. A partir de 1991 se inicia una etapa de expansión y crecimiento de la empresa, con la instalación de tiendas en los centros comerciales Parque Arauco y Plaza Vespucio. A su vez, a partir de dicho año, Almacenes Paris comienza a participar en el negocio inmobiliario de los centros comerciales, principalmente a través del grupo Plaza.

La empresa se abrió a la bolsa de valores a fines de 1996, colocando en el mercado bursátil 100 millones de acciones a un precio de \$350 cada una, recaudando aproximadamente US\$80 millones.

Durante el año 2003, se inician las gestiones para la apertura del Banco Paris a concretarse durante el año 2004.

Al 30 de junio de 2003, los activos consolidados de la empresa alcanzaban a US\$632 millones y las ventas en los últimos 12 meses llegaron a US\$548 millones.

Almacenes Paris maneja el negocio comercial de las tiendas por departamento, a través de las cuales está incorporando nuevos servicios como son el corretaje de seguros y agencia de viajes. Además, la empresa participa en el negocio inmobiliario de los centros comerciales, en el negocio financiero por el otorgamiento de

crédito a sus clientes, así como también en 2004 iniciará la operación del Banco Paris, y en la actividad industrial a través de la fabricación de muebles y derivados.

Almacenes Paris se ubica entre las principales tiendas por departamento del país, contando, actualmente con 16 de estos locales que tienen una superficie que varía entre los 4.000 m² y los 15.000 m². , 7 de los cuales se ubican en la Región Metropolitana.

Un formato de menor tamaño (entre 400 y 1.500 m²) ha comenzado a usar la empresa como una forma de atacar ciudades más pequeñas o de ser precursoras de una tienda completa una vez que ya se cuente con una base de clientes de tarjetas de crédito. A la fecha son dos las tiendas bajo este formato; Villa Alemana y Antofagasta.

Además, la empresa cuenta con el local "Centrofertas" en Santiago y dos tiendas Muebles Paris (La Serena y Talca).

La empresa divide sus líneas de negocios comerciales en Hogar y Vestuario, las que incluyen vestuario femenino, masculino e infantil, accesorios, artículos para deportes, electrodomésticos, cosméticos, artículos de perfumería, muebles, etc. Las ventas de la empresa se canalizan en un 55% a través de la línea hogar y en un 45% en vestuario.

D1.4. La Polar:

La Polar, adquirida en 1999 por el fondo estadounidense Southern Cross, es la cuarta cadena de tiendas por departamento en el país.

Participa a su vez en el negocio financiero a través de una tarjeta de crédito propia. A fines de 2005, La Polar contaba con 26 locales, con un total de superficie de ventas de 81 mil m², 1,4 millones de clientes activos e ingresos en 12 meses a marzo de 2006 por US\$487 millones. En 2006 La Polar agrega 4 nuevas tiendas y remodela otras 4 con lo cual llega a 128.235 M² con un tamaño promedio de 4.275m².

La Polar nace en Chile en el año 1920 y se desarrolla como una multitienda concentrada en la Región Metropolitana. En 1999, cuando la cadena contaba con 12 locales y una situación financiera deteriorada, es adquirida por el fondo de inversiones estadounidense Southern Cross Latin American private Equity Fund, a través de la Sociedad Comercial Siglo XXI, (ahora Empresas La Polar S.A.). Tras el cambio de propiedad se inicia un fuerte período de expansión para la compañía, con la apertura de nuevos locales, la expansión hacia regiones, el impulso del área financiera y el desarrollo de una estrategia de crecimiento a futuro.

La Polar se abrió a la Bolsa de Comercio en septiembre de 2003 recaudando cerca de MMUS\$32 equivalente a un 20% de la empresa. Luego, en noviembre, el controlador de la sociedad vendió un 26,1% de las acciones, equivalentes, en ese entonces, a MMUS\$52.

Finalmente, en enero de 2005 el Fondo Southern Cross redujo su participación a un 20,1%, con lo que el *free float* de la compañía se acerca a un 80%.

Hoy en día La Polar se consolida como la cuarta tienda departamental del país, comercializando vestuario y productos para el hogar, además de ofrecer financiamiento a sus clientes, a través, de una tarjeta de crédito propia.

La Polar seguirá enfocándose en el segmento socioeconómico C2C3D, en el cual posee un fuerte posicionamiento. El foco de la empresa seguirá siendo su área de retail, ya que, el negocio financiero (cada vez más relevante dentro del EBITDA consolidado) es visto como un área que se desprende de un buen funcionamiento del área de retail. En este sentido La Polar continuará su plan de expansión de locales, aumentando su presencia en malls y su cobertura a nivel nacional.

La empresa continuará explorando proyectos en común con otros operadores de retail, que le permitan crear polos de atracción para sus consumidores. En el aspecto financiero, la empresa continuará en el desarrollo de nuevos productos que le permitan continuar con su crecimiento.

La Polar participa en el comercio al detalle o al por menor a través de una cadena de tiendas por departamento en el país. Además, la empresa participa en el negocio financiero por medio del otorgamiento de crédito a sus clientes. Como es posible apreciar, la participación del negocio financiero como porcentaje de las ventas es bastante menor en relación a la contribución sobre el EBITDA de este mismo, en donde alcanza una participación del 77%.



Fig. D.3. Distribución de las ventas 2005 de La Polar, Fuente: Memoria Anual La Polar

D2. Las grandes tiendas en cifras

A partir de la información de los apartados anteriores y que corresponde a lo publicado por las propias empresas bajo estudio y empresas especializadas al igual que sus especialistas como [JR05], a continuación el siguiente cuadro presenta una visión global de la industria de las grandes tiendas con las siguientes cifras consolidadas:

Indicador	Falabella	Ripley	Paris	La Polar	Total
Nro. Locales	30	31	19	26	106
Superficie Venta m ²	177,538	227,909	154,544	81,000	640,991
Superficie / Locales	5,917	7,352	8,134	3,115	6,048
Ventas (MMUS\$)	1,178	871	782	360	3,194
% Mercado	36.90%	27.30%	24.50%	11.30%	100
Ventas/ m ²	3.55	2.6	2.6	2.3	
Tarjetas Emitidas (MM)	3.3	2.6	3.0	1.9	10.8
Tarjetas Activas (MM)	2.6	1.4	1.2	1.4	6.6
% Ventas con tarjetas	67%	63%	67%	80%	
Inversiones Proyectadas (MMUS\$)	1130	551	1200	100	

Tabla D.1. Cuadro Resumen de los indicadores macro de las tiendas líderes de la industria, fuente Adaptación de [JR05]

Las positivas tendencias de esta industria impulsan y proyectan inversiones entre 2006 y 2009 que superan USD\$ 3.000 millones.

D3. La industria de las grandes tiendas en Chile

El mercado de vestuario y hogar en Chile, operado a través del formato de tiendas por departamento, está más bien acotado y aunque se observa crecimiento a través de la apertura de nuevas tiendas, el gran dinamismo lo imprimen las fusiones de los grupos económicos subyacentes, esta actividad comercial es un sector industrial muy importante dentro del conjunto del sistema económico chileno, como se aprecia en los datos indicado en el apartado anterior y que se consignan en las publicaciones financieras como Jarufe [JR05], CCS [CC06], FitchRating [FR06a, FR06b, FR06c].

Según Jarufe [JR05], El área comercial de tiendas por departamento en Chile cuenta con alrededor de cien tiendas, alcanza tamaños de salas de venta que entre Falabella, Paris y Ripley se sitúan, en promedio, en los 6.800 m², mientras que, La Polar, bordea los 3.000 m².

El siguiente es un cuadro comparativo de las superficies y ubicaciones de las tiendas que conforman tres de las cuatro principales cadenas del país.

Ubicación	Almacenes Paris		Ripley		La Polar	
	Tienda	m2	Tienda	m2	Tiendas	M2
Antofagasta	Antofagasta	1,150	Antofagasta	12,033	ANTOFAGASTA	5,669
Calama	MallCalama	5,380	Calama	6,748		
Chillán	PlazaELRoble	5,180	Chillán	2,273		
Concepción	BarrosArana	12,359	Castellón	5,372	BIO - BIO	6,692
Concepción	PlazadelTrébol	8,470	ElTrébol	8,070	CONCEPCION	2,992
COPIAPÓ					COPIAPÓ	3,965
COQUIMBO					COQUIMBO	5,012
Curico					CURICO	2,552
EL BELLOTO					EL BELLOTO	3,906
Iquique			Iquique	7,620	IQUIQUE	5,537
LINARES					LINARES	4,734
LaSerena	PlazaLaSerena	5,860	LaSerena	7,061		
LosÁngeles	PlazaLosÁngeles	5,700			LOS ÁNGELES	3,354
OSORNO					OSORNO	5,428
PuertoMontt			PuertoMontt	6,794	PTO MONTT	6,004
QUILICURA					QUILICURA	4,445
RANCAGUA			Rancagua	10,659	RANCAGUA	4,207
Santiago Centro	Alameda	10,106	Agustinas	1,222	AHUMADA	2,960
Santiago Centro			Alameda	5,021	ALAMEDA	4,595
Santiago Centro			Astor	3,996	MONJITAS	1,047
Santiago Centro			Crillón	7,756	PUENTE	7,083
Santiago Centro			Huérfanos	3,951	SAN DIEGO	3,411
Santiago Centro			Puente	10,931		
Santiago Centro			SanDiego	4,496		
Santiago Conchali			PlazaNorte	7,254		
Santiago Cerrillos	PlazaOeste	9,394			COLINA	1,807
Santiago Colina						
Santiago LaFlorida	PlazaVespucio	10,519	FloridaCenter	14,017	LA FLORIDA	7,217
Santiago La Florida			PlazaVespucio	11,945		
Santiago LasCondes	AltoLasCondes	13,842	AltoLasCondes	11,902		
Santiago LasCondes	ParqueArauco	14,686	ParqueArauco	16,722		
Santiago Maipu			PlazaOeste	7,966	MAIPÚ	5,202
Santiago Ñuñoa			Irarrázaval	1,816		
Santiago Providencia	PlazaLyon	9,199	BarrosArana	2,178		
Santiago PuenteAlto	PlazaTobalaba	10,415	PlazaTobalaba	8,185	PUENTE ALTO	6,096
Santiago Renca					PANAMERICANA	1,538
Santiago San Bernardo					SAN BERNARDO	5,159
Santiago San Joaquín	CentroOferta	3,500				
Santiago San Miguel					GRAN AVENIDA	4,766
Santiago La Reina					LA REINA	3,765
Talca	Talca	3,988			TALCA	3,551
Temuco	Temuco	10,385	PortalTemuco	9,924		
Temuco			Temuco	5,987		
Valdivia			Valdivia	994	VALDIVIA	2,306

Valparaíso			Valparaíso	8,814		
VillaAlemana	VillaAlemana	1,200				
ViñadelMar	MarinaArauco	13,211	MarinaArauco	10,696	VIÑA DEL MAR	8,181
ViñadelMar			ViñadelMar	5506		

Tabla D.2. Distribución Regional en Chile de las Grandes Tiendas.

En 2004, estas cadenas totalizaron ventas cercanas a US\$2.600 millones, con participaciones de mercado de 36,6%, 25,6%, 28,0% y 9,8%, respectivamente. En el área financiera, con alrededor de 10 millones de tarjetas de crédito emitidas y 60% de ellas activas, las cuatro tiendas por departamento alcanzaron colocaciones brutas por \$843 mil millones a fines de 2004, cifra que representa un 13% de crecimiento con respecto al 2003. Las securitizaciones de carteras realizadas por La Polar (en 2002 y 2004) y Ripley (2002) agregan \$131 mil millones a la cartera total. A lo anterior se deberá agregar también la securitización que recientemente realizó Paris. Los niveles de morosidad de las carteras de colocaciones, medidos como la relación stock de provisiones / stock de colocaciones fluctúa ampliamente dependiendo del nivel de riesgo de los clientes situándose en 3% para Falabella, que presenta la cartera menos riesgosa y sobre 10% para carteras de segmentos socioeconómico menores como es el caso de La Polar.

Los principales protagonistas de esta actividad comercial en Chile son, en primer lugar, el comercio mayorista y, en segundo lugar, el comercio minorista. Tanto el comercio minorista como el mayorista son sectores de actividad económica relevantes en Chile, cuyo proceso evolutivo, se describe en el apartado anterior, y se resume en cuadro siguiente, señalando que Las tiendas por departamento en Chile tienen sus orígenes a fines del siglo XIX, con la creación de Falabella como una sastrería ubicada en el centro de Santiago. Una década después nace Almacenes Paris (Paris), originalmente como bajo el nombre de Muebles Paris. El desarrollo del negocio del crédito a través de una tarjeta de crédito propia es muy posterior y viene luego de que las empresas hubieran extendido sus líneas de productos incorporando electro-hogar. Incluso Falabella, ya había iniciado su expansión a regiones. Los últimos quince años han sido de fuerte crecimiento para las tres mayores cadenas de tiendas en el país; Falabella, Ripley y Almacenes Paris, mientras que La Polar ha acelerado su proceso de expansión los últimos cinco años, luego de que el grupo Southern Cross adquiriera la propiedad de la empresa

en 1999, la siguiente figura muestra en forma resumida esta evolución histórica:

Cronología de las Tiendas por Departamento en Chile

	Falabella	Aim. Paris	Ripley	La Polar
Inicios	1889	1900	1956	1920
Inicio Negocio de Crédito	1980	1980	1976	
Inicios en Regiones	1962 (Concepción)	1995 (Concepción)	1986 (Concepción)	1992 (Rancagua)
Ingreso a un mall	1983 (Parque Arauco)	1991 (Parque Arauco)	1993 (Parque Arauco)	2002 (Portal La Reina)
Inicia negocio inmobiliario	1990 (Plaza Vespucio)	1994 (Plaza Oeste)	1988 (Panorámico)	No
Salida al Exterior	1993 (Argentina)			
	1995 (Perú)	No	1997 (Perú)	No
Apertura a Bolsa de Valores	1996	1996	2005*	2003
Inicio Negocio Bancario	1998	2004	2002	No

* Ripley Corp. se abriría al mercado bursátil en julio de este año

Tabla D.3. Cronología de las tiendas por departamento en Chile, Fuente [JR05]

En la actualidad, el mercado del vestuario y hogar en Chile se caracteriza por el funcionamiento paralelo de dos sistemas: por una parte, un sistema basado en un comercio tradicional, formado por numerosos medianos y pequeños establecimientos, con estrategias de diferenciación, que actúan de forma independiente, que tienen una orientación específica a determinados segmentos del mercado. Esta situación de dualismo no equilibrado dentro de la distribución comercial provoca una competencia constante del comercio tradicional por el sistema formado por las grandes organizaciones con régimen de venta por departamento.

Este hecho se traduce en que el dominio del mercado del vestuario y hogar en Chile se encuentra en posesión de muy pocas compañías, fundamentalmente empresas que trabajan el formato de grandes tiendas de ventas por departamento. En este sentido, las primeras compañías del ranking de las tiendas de departamentos en Chile según se señala Jarufe en [JR05], estas cuatro principales empresas del sector; Falabella, Ripley, Almacenes Paris y La Polar, según se indicó alcanzaron ventas por US\$3.194 millones durante el año 2005.

Por tanto, este mercado del vestuario y hogar o la industria de las grandes tiendas de ventas por departamento se configura como un sector con una estructura de carácter oligopolístico con muy pocos grandes grupos dominando los diferentes formatos más relevantes

(grandes tiendas, tiendas medianas y tiendas pequeñas), y donde la dinámica competitiva es muy intensa.

Según Vázquez et al [VR97], la industria de las grandes tiendas, al encontrarse entre la producción y el consumo, crea *utilidades* a los consumidores y *servicios* a los productores. La distribución crea al consumidor utilidad de lugar, de tiempo, de forma y de creación de surtidos y de posesión:

1. Utilidad de lugar. Es creada por medio del transporte de los productos desde los lugares de producción hasta los de consumo, y también mediante la existencia de suficientes puntos de venta próximos al lugar donde el consumidor necesite el producto.
2. Utilidad de tiempo. Pone el producto disponible en el momento en que el consumidor desea consumirlo. Para ello, el distribuidor comercial deberá almacenar el producto en los almacenes o en las estanterías de los puntos de venta a la espera del momento en que el consumidor lo solicite, evitándole de este modo que tenga que comprar y guardar grandes cantidades de producto para su posterior consumo.
3. Utilidad de forma y de creación de surtidos. Adapta el producto comercializado a las necesidades de los consumidores. Por otra parte, la producción tiende y debe tender hacia la especialización para ser más eficiente, mientras que el consumo tiende a la complejidad, de tal forma que la distribución comercial creará surtidos de productos ajustados a las necesidades del consumidor para que pueda adquirirlos conjuntamente.
4. Utilidad de posesión. Finalmente, con la entrega de la cantidad de producto solicitada por el consumidor, contribuye a crear utilidad de posesión, ya que para que el producto genere utilidad al cliente es necesario que adquiera la propiedad o la posesión del mismo y pueda consumirlo.

Estas utilidades creadas en el mercado de vestuario y hogar para los consumidores tienen una relación directa con los servicios que crea

para los productores. Estos servicios serían de transporte, almacenamiento, "finalización del producto", información, financiación y asunción de riesgos:

1. *Servicios de transporte.* Transporta los productos desde la fábrica hasta los puntos de venta. Esto se debe a que las actividades productivas se encuentran concentradas en localizaciones donde pueden obtener mejores ventajas de oportunidad (por ejemplo, en mano de obra, materias primas, tecnología, terrenos, etc.) mientras que los consumidores se encuentran concentrados en grandes ciudades o en pequeñísimas poblaciones.
2. *Servicios de almacenamiento.* Almacena los productos después de que salen de la cadena de fabricación, y por tanto, el productor no debe preocuparse por problemas de almacenamiento de stocks. Además, se debe tener en cuenta que el ritmo entre producción y consumo es distinto: la producción sigue un ritmo discreto, es decir, se produce cuando las condiciones son más ventajosas, mientras que el ritmo del consumo es continuo, de tal forma que la distribución se encargará de almacenar los productos para cuando el consumidor los solicite.
3. *Servicios de "finalización del producto".* Estos servicios incluyen el fraccionamiento, la clasificación, la normalización, la presentación (envase, empaquetamiento), etc., y son creados también por la distribución comercial.
4. *Servicios de información.* La distribución comercial informa al productor o fabricante sobre determinados aspectos relevantes, como pueden ser precios, calidades, cantidades, establecimiento de contactos, tendencias en la moda, etc. Esta información es muy valiosa para el productor, ya que la distribución comercial está en contacto constante con el consumidor conociendo de manera más próxima sus necesidades.
5. *Servicios de financiación.* Componente importante de esta actividad económica es el financiamiento del fabricante debido a que compra sus productos se pagan a plazo por ejemplo a 30 días y el distribuidor tarda en venderlos más tiempo. En la actualidad, esta situación es

contraria para el caso de las grandes empresas de distribución, ya que pagan a los productores a un plazo de tiempo superior al que ellos tardan en vender esos productos, por lo que es el fabricante el que financia a los distribuidores.

6. *Asunción de riesgos.* La distribución comercial asume riesgos al comprar los productos al fabricante, ya que puede ocurrir que estos no se vendan porque queden obsoletos, se estropeen en el almacén, los roben, etc. Estos riesgos no los corre el fabricante desde el momento que pone sus productos a disposición de la distribución comercial. Generalmente, estos riesgos suelen ser cubiertos por la distribución comercial a través del pago de pólizas de seguros, también con acuerdos comerciales para corregir tales situaciones como liquidaciones o ventas a saldos.

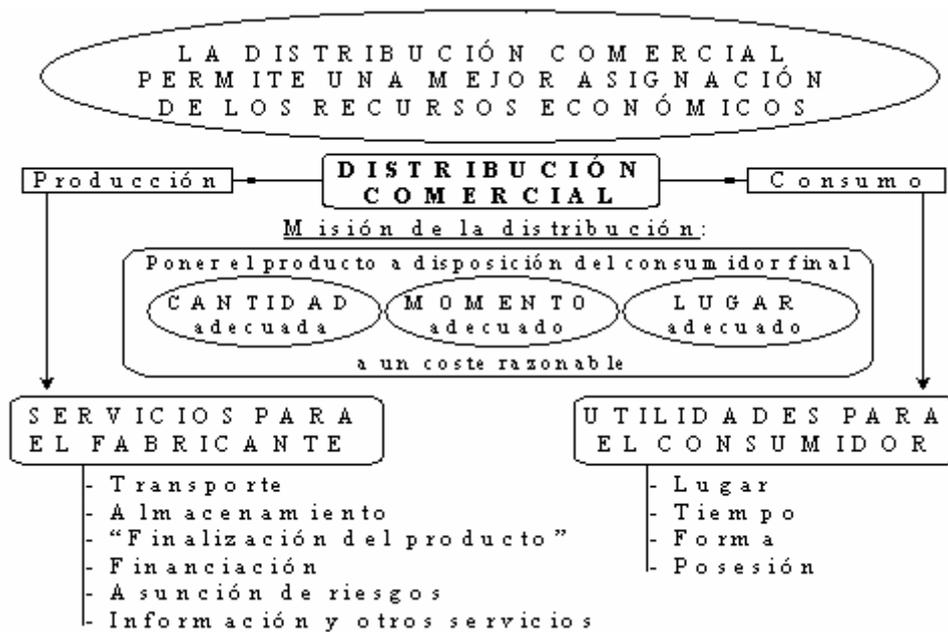


Fig. D.5. Función o Instrumentos del Marketing, fuente: [VR97]

En definitiva, la distribución de vestuario y electro-domésticos es el puente que une la producción con el consumo debido a que ofrece al consumidor "el *producto adecuado*, en el *lugar adecuado* y en el *momento adecuado*" y a un costo razonable. Las utilidades que son añadidas a los productos para los consumidores, así como los servicios ofrecidos a los productores suponen un coste que elevará el precio de

venta de los productos. Este coste deberá ser razonable y acorde con las utilidades y servicios creados para que el consumidor esté dispuesto a pagarlo.

Según lo indicado por Santesmases en [SM99], desde el punto de vista macroeconómico, la distribución de vestuario y electro-doméstico permite además una mejor asignación de los recursos económicos al poder especializarse la producción por zonas geográficas, en función de los recursos naturales de las mismas y de la capacidad y formación de sus habitantes facilitando la distribución posteriormente que los excedentes puedan ser intercambiados entre zonas geográficas por otros productos que se necesitan y no se disponen. Por tanto, la distribución comercial de vestuario y hogar tiene la función económica de facilitar los intercambios de bienes y servicios reduciendo costes al poner en contacto la producción y el consumo, de tal forma que favorece el desarrollo industrial y satisface las necesidades de compra de la población.

Casares et al [CJ96a], señala que la *relación fabricante-distribuidor* se ha caracterizado en las últimas décadas por la pugna por ser "capitanes del canal". Esta pugna ha tenido como efecto la entrega del poder por parte del fabricante al distribuidor que posee un elevado volumen de compras y que ofrece un número reducido de referencias expuestas en sus estanterías de venta. La aparición de grandes empresas de distribución ha significado que éstas pueden dar salida al total de la producción de muchos proveedores, a la vez que estas empresas de distribución son necesarias para mantener los productos en el mercado, sin condenarlos a una presencia marginal o en ciertos mercados locales, sin posibilidad de expansión. En este sentido, Lamo de Espinoza et al [LJ94], señala que en esta situación, el distribuidor ostenta un indudable poder a la hora de negociar sus compras con el fabricante debido a las diferencias en la participación de ventas totales (el porcentaje sobre las ventas totales de una empresa importante de distribución de los productos de un determinado fabricante es mínimo, mientras que las ventas realizadas por el fabricante a estas empresas de distribución pueden suponer un porcentaje notable sobre el total. Cruz Roche [CI 97], agrega que en la negociación, no sólo se busca obtener unos menores

precios de compra, sino también conseguir una financiación de acciones promocionales (folletos, publicidad, cabeceras de góndola, etc.), descuentos por servicios (reposición), descuentos por pago centralizado, apertura de nuevos centros, altas de nuevas referencias, etc., y, finalmente, condiciones de aplazamiento de pagos.

D3.1. Tendencias generales de la venta por departamento

La tendencia de esta actividad económica en Chile ha sido la *internacionalización*, en la actualidad, las oportunidades de negocio para las grandes cadenas de ventas por departamento en los mercados nacionales son cada vez más reducidas debido a los altos niveles de competencia y a las similitudes de sus formatos comerciales que compiten por el gasto de los mismos consumidores. Como consecuencia, las grandes empresas de venta al detalle se han introducido en nuevos mercados extranjeros donde la competencia no es tan fuerte y así poder incrementar su cifra de beneficios. Además, esta tendencia de internacionalización de estas empresas se desarrolla también para conseguir un mayor poder de compra frente a mayoristas y fabricantes y para poder conseguir economías de escala. Esta tendencia comienza en Chile en 1993 con Falabella en Argentina, en 1995 con Ripley en Perú.

Otra tendencia en el proceso de internacionalización de esta actividad, fue el *crecimiento externo* de estas empresas mediante la compra de otras empresas o la fusión con ellas, como lo es el caso de Cencosud en Argentina y las primeras incursiones de Ripley y Falabella en Perú.

D3.2. Otras tendencias generales de la distribución comercial

Además de las actuales tendencias en Chile y de las tendencias generales del sector retail en el mundo globalizado, existen otras más específicas que deben tenerse en cuenta, como las que plantean Kotler en [KP00] y [KP99] y Vázquez en [VR97], las cuales se mencionan a continuación:

1. El *ciclo de vida de todos los formatos comerciales se acorta*, alcanzándose antes su madurez. Por ello, unos

formatos sucederán a otros con mayor rapidez. Esto se debe a que cuando un formato comercial novedoso tiene éxito, rápidamente es imitado por otros, haciéndole perder enseguida ese carácter novedoso [KP00].

2. La *integración vertical* de los canales de distribución es un hecho creciente. Existe una tendencia hacia el acortamiento de los canales de distribución, en busca de alcanzar un control más efectivo de las funciones de distribución. Los canales de distribución excesivamente largos son ineficientes debido a los mayores costes añadidos en los controles y a la dificultad de motivar a una larga red de intermediarios independientes [VR97].

3. Se está produciendo paulatinamente un *aumento del poder de negociación de las grandes cadenas de distribución* frente a los fabricantes. Gracias a que tienen sistemas de información superiores pueden utilizar información de marketing sofisticada y sistemas logísticos para suministrar un buen servicio e inmensos volúmenes de producto a precios muy atractivos para masas de consumidores. En este proceso, la gran distribución está excluyendo a los fabricantes más pequeños, que se ven dependientes de una gran empresa de distribución comercial y son, por tanto, extremadamente vulnerables. Muchas empresas de distribución comercial están incluso diciendo a los fabricantes más poderosos lo que tienen que hacer, cómo deben fijar sus precios, cómo deben hacer la promoción de sus productos, cuándo y dónde expedir la mercancía, e incluso cómo reorganizar y mejorar la producción y la gestión. En esta situación los fabricantes tienen pocas alternativas, ya que oponerse a las directrices de la distribución comercial podría

suponer arriesgar entre un 10 y un 30 por ciento del mercado [KP00].

4. Para una eficiente gestión de la distribución será necesario la *incorporación y difusión de nuevos y modernos sistemas tecnológicos* (el escáner, el etiquetado electrónico, el E.C.R., el intercambio electrónico de datos, etc.).
5. El *merchandising será un factor diferenciador*. El consumidor, que es cada vez más exigente, busca cada vez más comprar en un ambiente agradable, lúdico y sofisticado.
6. Los *fabricantes se transforman en distribuidores*. Son cada vez más numerosas los fabricantes que abren espacios de venta con su propio nombre con el fin de optimizar su marca. Un ejemplo de esta situación en Chile son las tiendas "factory" (Nike, Levi's, etc.).
7. Cada vez más empresas adoptan el concepto de *gestión integrada de la logística*, que se basa en una concepción de la distribución física que reconoce la necesidad que tiene la empresa de integrar su sistema de logística con los de sus proveedores y clientes con el objetivo de maximizar los resultados de todo el sistema de distribución [KP99].
8. La *demanda está cada vez más segmentada*, por lo que existirán muchas oportunidades para aquellas empresas que desarrollen una estrategia de nicho o de especialista.
9. Finalmente, la *era electrónica* ha acelerado de forma significativa el crecimiento de la venta sin establecimiento. Los consumidores reciben ofertas a

través de sus televisores, ordenadores y teléfonos, a las que pueden responder de forma inmediata a través de llamadas gratuitas telefónicas o del propio computador [KP00]. Por tanto, la tendencia de la venta a través de nuevos canales es más que notable. Entre estos nuevos canales hay que destacar la gran importancia del comercio electrónico.

D3.3. Clasificación de los formatos comerciales tradicional y actual

Según Vázquez [VR97], el comercio tradicional se caracteriza por ser necesaria la presencia del comerciante que presenta y entrega el producto solicitado al comprador, ya que los productos se localizan normalmente fuera del alcance de las manos de éste, pues se encuentran situados principalmente tras el mostrador. En esta situación, el comerciante desempeña un papel crucial desde una doble perspectiva: por un lado, el comerciante presenta las argumentaciones y demostraciones necesarias para resaltar las características de los productos y actúa como prescriptor en el momento de aconsejar qué producto o marca debe adquirir el comprador, y por otro lado, el comerciante puede transformar los deseos del comprador en actos de compra, debido a que se establece una relación con el cliente, el cuál le plantea sus necesidades al comerciante y éste procura satisfacerlas. Dentro del comercio tradicional de productos de gran consumo se pueden distinguir dos tipos de establecimientos: *tiendas clásicas o tradicionales* y *tiendas especializadas*.

La tienda clásica o tradicional es aquella que ofrece un surtido amplio pero poco profundo debido a la pequeña dimensión de sus establecimientos, como por ejemplo los clásicos establecimientos de especialidad. Este tipo de formatos comerciales de productos de gran consumo, que no disponen de servicios adicionales para atraer a los clientes, no pueden competir con otros formatos más evolucionados, fundamentalmente basados en el sistema de venta del libre servicio, por lo que se encuentran en una situación de franco retroceso. Debido a este hecho, en muchos casos, estas tiendas tradicionales se han reconvertido hacia formas más evolucionadas como el autoservicio o se han introducido en cadenas corporativas.

Por su parte, la *tienda especializada* es aquella que ofrece un surtido muy reducido pero profundo en sus pequeños establecimientos, como por ejemplo, una tienda de electrónica, o electrodomesticos o artículos deportivos, una lencería, una camisería, una sastrería, etc. Precisamente, el arma competitiva de este tipo de establecimientos es la selección, variedad y calidad de un surtido idóneo para un comprador que no sólo tiene en cuenta las ofertas en precios a la hora de elegir el establecimiento de compra. Aunque este tipo de formato es menos vulnerable a la competencia de los formatos basados en el autoservicio, fundamentalmente porque trabajan con líneas de productos difícilmente integrables en el sistema de libre servicio, se están viendo también afectados por la atracción que ejercen las grandes tiendas.

D3.4. Formato de autoservicio

Para Masa [ML93], el *autoservicio* es la tecnología de venta que ha supuesto mayores transformaciones en el comercio de nuestros días debido a que la relación comerciante-cliente es vista desde una óptica radicalmente distinta a la tradicional. En el sistema de autoservicio es el cliente el que toma directamente los productos de los expositores o estanterías, los examina, los compara, los agrupa y los transporta hasta la salida del establecimiento donde abona el importe total de su compra en las cajas de salida, donde existe personal exclusivamente dedicado al cobro, siendo éste, salvo raras excepciones, el único contacto con el personal del establecimiento. Con la introducción del sistema de autoservicio, la función de prescriptor del comerciante desaparece, provocando esta eliminación del principal papel del comerciante la aparición de la palabra distribuidor, inventada por el francés Edouard Leclerc, que rebaja al comerciante al rango de repartidor (Miquel Peris *et al* [MS96]). Para Casares [CJ97], esta tecnología de venta presenta, tanto para el comprador como para el comerciante, una serie de ventajas: para el comprador, comodidad y servicio rápido (ahorro de tiempo), libertad de elección y movimientos en las decisiones de compra, comparar precios y marcas y posibilidad de conseguir precios más bajos (si se trasvasan a los mismos las

economías de coste); y para el comerciante, racionaliza la exposición de productos fomentando la venta impulsiva a través de técnicas de merchandising, aumenta la productividad por persona empleada al ahorrar mano de obra, posibilidad de ofrecer una mayor gama de productos y de aumentar la productividad por metro cuadrado como consecuencia de sustituir espacio de almacén por espacio de venta.

En Chile, el concepto de autoservicio se introduce junto con la ampliación de la oferta en las tiendas de Falabella durante los años 50, desde entonces y hasta nuestros días, la fórmula del autoservicio ha seguido una evolución creciente, tanto por el número de establecimientos, como por la cuota de mercado captada por esta fórmula comercial.

Dentro de esta tecnología de venta, que se ha extendido notablemente en las últimas décadas, pueden distinguirse diferentes formas comerciales en el sector de productos de gran consumo, como son los que presenta ACNielsen en [AC05], con formatos tales como el *autoservicio*, *superservicio*, *supermercado*, *hipermercado* los *establecimiento descuento*, *tiendas de oportunidad*, *todo lo cual se consolida con los Power Center*, que reúne *Hipermercados más tiendas anclas o diversos comercios orientados a los servicios*.

Para Casares [CJ97] y Vázquez [VR97], el *autoservicio* se caracteriza por tener una superficie pequeña, que suele oscilar entre 40 y 99 metros cuadrados, por trabajar con una sola caja registradora, según la clasificación de Nielsen, y por ofrecer un surtido reducido. Muchos de los autoservicios existentes actualmente tienen su origen en una tienda tradicional, que ha evolucionado para adaptarse a las nuevas tendencias y a los nuevos hábitos de compra del consumidor, pero manteniéndose las restantes características del establecimiento tradicional en lo relativo a la política comercial, el servicio al cliente, la gestión rudimentaria y la escasa complejidad de las técnicas de merchandising utilizadas. Por tanto, en la evolución de los autoservicios se observa un claro dualismo: por un lado, el autoservicio que proviene de una transformación de un establecimiento tradicional y que se encuentra con los mismos problemas que ellos, y por otro, los que están integrados en cadenas

corporativas que pueden operar con una relación de precios y costes más favorables.

Miquel Paris [MS96], señala que el *superservicio* es un establecimiento minorista que vende en régimen de autoservicio y cuya superficie de venta oscila entre 100 y 399 metros cuadrados. La diferencia fundamental con el autoservicio, a parte del tamaño, es la existencia de secciones atendidas por personal especializado que realiza las funciones propias del despacho directo. Todo ello conlleva una concepción empresarial distinta, ya que se requiere la delegación de funciones, el desarrollo de un plan contable, la confección de un organigrama para la adecuada coordinación de las secciones, etc..

Una tienda de departamentos mediana (entre pequeña y grande) es un comercio con una superficie que oscila entre 400 y 2.499 metros cuadrados. Se pueden clasificar en pequeños (de 400 a 999 metros cuadrados) y grandes (de 1.000 a 2.499 metros cuadrados). La tienda pequeña ofrece un surtido de productos muy completos, tanto en calidad como en precios, para poder servir a diferentes segmentos de mercado. Además del vestuario y electro domésticos, este tipo de establecimientos ofrecen otras categorías de productos (perfumería, juguetería, etc.) ampliando así el surtido ofrecido a la clientela.

Según Muñiz et al [MN98], las causas que inducen a las empresas de distribución a crecer pueden ser: en primer lugar, la búsqueda de rentabilidades atractivas y, en segundo lugar, la reducción riesgos estratégicos. En este sentido, la búsqueda de rentabilidades atractivas del capital invertido puede ser la causa del crecimiento mediante la introducción en otros formatos comerciales, sobre todo por empresas de distribución comercial situadas en sectores maduros o en mercados saturados, como por ejemplo, la introducción de las principales empresas de venta por departamentos en el formato de tiendas anclas en los Mall o en las Power Center y tienda descuento o liquidadoras. Por otra parte, una empresa de distribución comercial altamente rentable puede plantearse una estrategia de crecimiento debido a que estaría expuesta a un grave riesgo estratégico, ya que las empresas competidoras actuales o potenciales podrían en un

futuro conseguir posiciones competitivas difícilmente atacables, lo que conllevaría la pérdida de esas altas rentabilidades.

Para Labin [LJ95], en definitiva, si una empresa de distribución comercial quiere perdurar competitivamente en el mercado debe plantearse de forma constante el desarrollo de estrategias de crecimiento, pero no sólo para mejorar las ventas, la cuota de mercado, el beneficio o el tamaño de la organización, sino también para sobrevivir a los ataques de la competencia, gracias a las economías de escala y a los efectos experiencia que ofrece.

Otro factor importante como lo señala Muniz en [MN95], en el planteamiento de llevar a cabo una estrategia de crecimiento también van a influir de forma importante las estructuras de propiedad de las compañías: en las empresas donde existe una separación entre la propiedad y la gestión, los intereses de los accionistas pueden entrar en conflicto con los de los directivos, debido a que los accionistas buscan un beneficio a corto plazo y las estrategias de crecimiento implican un riesgo que no están dispuestos a asumir, lo que condiciona e incluso impide el crecimiento de la sociedad. Este hecho no ocurre en las empresas donde la propiedad está más ligada a la gestión, ya que la planificación puede realizarse a más largo plazo, cuando es admitida por la empresa de distribución comercial la necesidad de desarrollar estrategias de crecimiento, la clave del éxito consiste en minimizar el riesgo que todo alejamiento del negocio básico implica maximizar la ventaja competitiva de la que dispone la empresa con respecto a otros operadores establecidos en el nuevo sector o mercado en el cuál quiere introducirse.

Las estrategias de crecimiento y expansión han sido estudiadas para las empresas de distribución comercial por algunos autores como por ejemplo Robinson et al [RT90], adaptando el modelo de Ansoff [AI76]. Desde este punto de vista, se pueden distinguir cuatro estrategias básicas de crecimiento y expansión para una empresa de distribución comercial:

1. *Estrategia de penetración en el mercado*: explotación del mismo formato comercial en el mismo mercado.

2. *Estrategia de internacionalización*: apertura a otros mercados geográficos con el mismo formato comercial.
3. *Estrategia de integración vertical*: extensión de las actividades de la compañía hacia actividades mayoristas y de producción.
4. *Estrategia de diversificación*: entrada en otros formatos comerciales y en sectores de apoyo a la actividad comercial.

Estas cuatro estrategias no suelen utilizarse de forma exclusiva, sino que las empresas de distribución comercial suelen aplicarlas de manera combinada. Además, estas estrategias se pueden desarrollar a través de dos alternativas: crecimiento interno u orgánico y crecimiento externo.

A continuación se revisan las diferentes alternativas de desarrollo (crecimiento interno u orgánico y crecimiento externo) y cada una de las estrategias de crecimiento o expansión planteadas por autores como Lambin y Kotler ya citados (penetración en el mercado, internacionalización, integración vertical y diversificación comercial).

Las empresas de distribución comercial o venta al detalle como las grandes tiendas pueden plantearse cuatro estrategias de crecimiento básicas:

1. *Estrategia de penetración en el mercado propio*. Según Kotler, consiste en el aumento de la participación en los mercados en los que opera y con el mismo formato comercial, pudiendo existir tres caminos para desarrollar esta estrategia: en primer lugar, que los clientes actuales consuman más productos y servicios; en segundo lugar, atraer clientes de los competidores; y en tercer lugar, atraer a clientes potenciales que no compran en la actualidad en este formato comercial. Esta estrategia puede desarrollarse a través de un crecimiento interno (por ejemplo, aumentando el número de tiendas propias) o a través de un crecimiento

externo (por ejemplo, a través de la compra o fusión de empresas competidoras).

2. *Estrategia de internacionalización.* La empresa de distribución comercial a través de esta estrategia se introduce en otros mercados geográficos con el mismo formato comercial. Esta estrategia presenta mayores niveles de riesgo cuanto mayor es la diferencia del mercado objetivo con los mercados de origen en cuando a estilos de vida, lenguaje, entorno cultural, requerimientos legales, renta per cápita, etc. En este sentido se puede hablar de dos niveles de internacionalización: en primer lugar, entrada en mercados con una afinidad socio-cultural y legal elevada; y en segundo lugar, entrada en mercados donde la afinidad es reducida. Para el caso chileno, los mercados del primer nivel estarían compuestos principalmente por los países del cono sur de América Latina. Esta estrategia de internacionalización se puede llevar a cabo por tres caminos: realización de una inversión directa (creación de establecimientos propios o compra de una empresa de distribución comercial local), realización de una *joint-venture* (creación de una nueva empresa con la asociación de una empresa local que aporte el conocimiento del mercado) y exportación del formato comercial a través de la fórmula de la franquicia.
3. *Estrategia de integración vertical.* Tiene lugar cuando la empresa de distribución comercial extiende su negocio hacia actividades mayoristas y de fabricación, pudiendo mejorar los precios y los servicios a sus clientes. En función de la dirección de la integración se puede hablar de integración hacia atrás (el minorista se integra con el mayorista o el fabricante) o integración hacia delante (el mayorista se integra con el minorista). Para el desarrollo de la estrategia de integración se pueden seguir tres vías: la integración corporativa, la integración contractual y la integración administrada.

4. *Estrategia de diversificación comercial.* Según Lambin, es la estrategia de crecimiento más arriesgada y más compleja, ya que conduce a la empresa a terrenos completamente nuevos para ella tanto en mercados como en productos, por lo que esta estrategia se justifica si el sector en el que se encuentra la empresa no presenta ninguna o muy pocas oportunidades de crecimiento o de rentabilidad, ya sea porque la competencia ocupa una posición demasiado fuerte, ya sea porque el mercado de referencia está en declive. Kotler plantea que esta estrategia implica la introducción paulatina de la empresa en otros formatos comerciales y en sectores de apoyo a la actividad comercial. En este sentido, se puede hablar de tres posibilidades para desarrollar la estrategia de diversificación: en primer lugar, la introducción en nuevos formatos comerciales que tengan sinergias tecnológicas o de marketing con los formatos comerciales existentes, aunque estos nuevos formatos comerciales se dirijan a un nuevo segmento de consumidores (estrategia de diversificación concéntrica); en segundo lugar, la explotación de nuevos formatos comerciales que no tengan sinergias tecnológicas con los formatos comerciales existentes, pero que se dirijan a los clientes actuales (estrategia de diversificación horizontal); y en tercer lugar, la introducción en nuevos negocios que no guarden relación ni con el formato comercial actual ni con los clientes actuales (estrategia de diversificación en conglomerado). Al igual que en las estrategias de crecimiento anteriores, la estrategia de diversificación se puede poner en marcha a través de un crecimiento interno o a través de un crecimiento externo.

D3.5. Las estrategias comerciales

La estrategia de penetración del mercado es la estrategia más utilizada por las empresas de distribución comercial. Esto se debe a que es la estrategia que conlleva un menor riesgo, ya que supone el desarrollo de formatos comerciales similares en el mismo mercado, es decir el desarrollo del negocio básico, sobre el que se posee un alto conocimiento, esta estrategia la ha llevado a cabo cada una de

las empresas bajo estudio con el propósito de cubrir los segmentos de mercado objetivo, en particular en la Región Metropolitana y en las principales Capitales Regionales. La de internacionalización es la alternativa de crecimiento por la que optaron Falabella y Ripley, las cuales tienen en común haber evolucionado desde la fabricación de confecciones a tiendas de venta por departamento. En esa evolución esta implícita la integración vertical como estrategia adoptada por estas empresas del sector textil, y en su momento también lo hizo Almacenes París que migro de la manufactura de muebles a las grandes tiendas, consiguiendo dominar por completo el canal de distribución y ofrecer mejores precios y mayor número de servicios.

D3.6. Estrategia de penetración en el mercado

La estrategia de penetración en el mercado consiste en incrementar la participación de la empresa de distribución comercial en los mercados en los que opera y con los productos actuales, es decir, en el desarrollo del negocio básico. Esta estrategia se puede llevar a cabo provocando que los clientes actuales compren más productos (por ejemplo, ampliando los horarios comerciales), atrayendo a los clientes de la competencia (por ejemplo, bajando precios) o atrayendo a clientes potenciales (por ejemplo, ofreciendo parking gratuito). Desde este punto de vista de las empresas de distribución comercial, esta estrategia consiste en crecer sobre la base del mismo formato comercial dirigido al mismo mercado, pudiendo desarrollarse esta estrategia o bien a través de un crecimiento interno, o bien a través de un crecimiento externo. El crecimiento interno se desarrollará a través de la apertura de nuevos establecimientos propios con el mismo formato comercial en el mismo mercado en el que opera la empresa, como ha sido las recientes inauguraciones realizadas por las cuatro grandes cadenas bajo estudio. Por su parte, la alternativa de crecimiento externo puede llevarse a cabo a través de dos posibilidades: en primer lugar, la compra de empresas de distribución que operan con el mismo formato comercial en el mismo mercado, como por ejemplo los situados en otros segmentos del retail como ha sido, la fusión de las grandes ferreterías y tiendas de mejoramiento del hogar de SODIMAC con Home Depot de Falabella. En cuenta Supermercados la compra de San Francisco por Falabella y

en sentido contrario la compra de Almacenes Paris por parte de Cencosud propietaria de los Hipermercados JUMBO, la participación de Fallabella en Farmacias Ahumada, entre otras.

Esta es, por tanto, una estrategia de crecimiento con reducido riesgo. Debido a esta circunstancia, esta debe ser una estrategia por la que deben optar en primer lugar las empresas de distribución comercial, ya que supone la explotación del negocio básico con un riesgo muy controlado debido al alto conocimiento del producto-mercado en el que trabaja la empresa.

La estrategia de penetración del mercado se puede combinar simultáneamente con otras estrategias de crecimiento más agresivas y, por tanto, más arriesgadas, como puede ser la internacionalización o desarrollo del mercado (introducción en otros mercados con el mismo formato comercial), el desarrollo de nuevos formatos comerciales (ofrecer nuevos formatos al mismo mercado) y la diversificación (desarrollo simultáneo de nuevos formatos en nuevos mercados).

D3.7. Estrategia de internacionalización

Las empresas de distribución comercial tradicionalmente han tenido un ámbito de actuación circunscrito a entornos geográficos locales, regionales o nacionales, debido en gran parte a las elevadas inversiones necesarias y al alto riesgo implícito para la expansión a otros países. Sin embargo, existen una serie de ventajas que empujan a las empresas de distribución comercial a optar por la internacionalización y a considerarla como una buena estrategia de crecimiento. En este sentido, Muñiz [MN95], señala las ventajas que pueden conseguir las empresas de distribución comercial desarrollando una estrategia de internacionalización son las siguientes:

1. La obtención de un suministro centralizado. Puede ser una ventaja competitiva para aquellas empresas que operan en mercados globales sobre las que operan en

2. La introducción de una nueva fórmula comercial o una mejora de una fórmula ya existente en un nuevo mercado. Esta innovación puede ser capitalizada acelerando la implantación de la nueva fórmula en el extranjero, ante la imposibilidad de proteger ese "saber hacer" mediante activos protegidos legalmente (patentes) como en las empresas manufactureras. Esta circunstancia explicaría el desarrollo acelerado en los países limítrofes donde aplicaron sus fórmulas comerciales Ripley y Falabella en Perú.
3. La obtención de economías de escala derivadas de una mayor dimensión empresarial. Esta es, a la vez, una ventaja para internacionalizarse y un objetivo cuyo fin es la búsqueda del aumento de la capacidad de compra desarrollando una red de distribución internacional que propicie economías de escala en el aprovisionamiento, aunque este objetivo puede conseguirlo una empresa de distribución comercial adhiriéndose a una alianza nacional o internacional de compañías de distribución, como es el caso de los supermercados Jumbo en Argentina y Tottus en Chile y las tres grandes cadenas de multitiendas chilenas en el extranjero.
4. La búsqueda de ventajas de localización. Se conseguirán mayores ventajas de localización en aquellos países donde la afinidad socio-cultural con el país de origen del formato comercial sea mayor. En

5. Necesidad de crecimiento de empresas distribuidoras que se enfrentan en sus respectivos países a mercados maduros en los que resulta difícil crecer con su actividad principal, es el caso de Chile para sus principales cadenas de multitiendas.
6. Aprovechamiento de oportunidades de negocio en otros países. Como fue lo ocurrido con Argentina y Perú para Falabella y Ripley, que decidieron incursionar

en los países vecinos porque vieron muy buenas oportunidades de negocio para asentar sus actividades, aunque en Chile tenían y tienen un mercado con una capacidad de negocio sostenido.

7. Adquisición de experiencia asociándose con empresas ya establecidas. Esto se debe a que el negocio de la distribución comercial requiere de dos tipos principales de saber hacer: el conocimiento del mercado y del consumidor local, y el dominio de las técnicas de gestión más eficientes y efectivas, estrategia utilizada por Cencosud en Argentina, Falabella con SAGA en Peru y Corona en Colombia.

8. Otros factores. Caben destacar la disminución y eventual desaparición de las barreras al flujo de capitales, la promoción institucional de proyectos multinacionales a nivel latinoamericano con tratados de integración, las nuevas tecnologías en la comunicación y mejoras en las por ejemplo, la no existencia de fuertes competidores, mayor será la aplicación de estrategias de internacionalización) y las limitaciones legales tanto del país infraestructuras, la presencia de cuadros directivos más profesionales y con una mayor disposición a establecerse en otros países, etc.

Existen una serie de factores que influyen en la decisión de llevar a cabo una estrategia de internacionalización por una empresa de distribución comercial. Estos factores, como lo señala Casares et al [CJ91], se pueden clasificar en dos tipos: factores impulsores y factores de freno.

1. Los factores impulsores. Son las oportunidades percibidas en los países de destino que suponen un atractivo para los grandes distribuidores: mercados poco desarrollados, posibilidad de obtener "nichos de mercado", posibilidad de desarrollar una "cabeza de puente" para la posterior

expansión, aprovechamiento del "saber hacer" de la empresa matriz, etc.

2. Los factores de freno. Consisten en situaciones del país de origen que empujan a las grandes empresas a salir al exterior, como pueden ser: mercados maduros o saturados, presión de la competencia, restricciones legales, estancamiento de la población, etc.

Una vez decidida la internacionalización por una empresa de distribución comercial ésta puede seguir dos posibilidades genéricas:

1. Exportar su concepto. Reproducir el formato comercial más o menos de forma idéntica, es decir, explotar la experiencia adquirida en el país de origen, como fue el resultado obtenido por Ripley y Falabella en Perú. Esta posibilidad permite a la empresa beneficiarse de economías de escala a nivel de producción, de distribución y de comunicación, pero deberá tener en cuenta los límites de la estandarización y la necesaria adaptación a las circunstancias locales.
2. Adquirir totalmente o en parte una empresa existente en el país de destino, Como lo realizó CencoSud con Disco en Argentina. Esta estrategia, al margen del reparto de riesgos financieros o de la investigación de tasas de rentabilidad superiores a las del país de origen, permite a la empresa distribuidora la implantación de establecimientos en un país donde la empresa no dispone de una ventaja competitiva suficiente para batir la competencia local.

Los posibles caminos para la internacionalización según lo plantea Torjaman et al [TA91], para penetrar en mercados extranjeros por parte de las empresas de distribución comercial, las vías o estrategias de internacionalización pueden concretarse con las cuatro siguientes estrategias:

1. Apertura directa de puntos de venta (inversión directa). Supone la existencia de barreras de entrada superables y

2. Adquisición de cadenas nacionales. Consiste en la compra total o parcial de una empresa distribuidora comercial asentada en el país objetivo. Frente a la inversión directa tiene la ventaja de partir de una red experimentada con un conocimiento del mercado aprovechable, estrategia implementado por Cencosud en Argentina.

3. Creación de negocios conjuntos (joint-ventures). Según explica Santesmases en [SM99], consiste en introducirse en un mercado extranjero asociándose con una empresa local, por lo que se crea una nueva empresa, cuyo capital es compartido. Esta asociación puede llevarse a cabo mediante la creación de una sociedad, en la que cada empresa invierte una parte del capital, o bien sólo con un contrato, en el que se establecen las funciones, aportaciones, participaciones y distribución de beneficios. Por tanto, una joint-venture consiste en una nueva sociedad subsidiaria creada por dos o más sociedades matrices, con personalidad jurídica independiente, pero supeditada a las estrategias competitivas de las empresas propietarias. En esta situación, la empresa exportadora ofrece su nueva fórmula comercial y su "saber hacer" y la empresa local ofrece su conocimiento sobre el mercado. Así fue como Ripley inicio la apertura de Grandes Tiendas en Perú, cuando en 1997 inaugura su primera tienda fuera de Chile, Ripley Jockey Plaza, en Lima Perú y en 1999, inicia la operación de la Financiera Cordillera (Financor) en Perú, la que fue creada para ofrecer tarjetas de crédito a los clientes de Ripley. Por tanto, exporta la experiencia o la fórmula en conjunto y por motivos de diversificación de riesgos, de búsqueda de habilidades complementarias o de

restricciones legales, se da entrada a una parte de capital del país de destino. Similar experiencia esta realizando Falabella con la Familia Corona en Colombia.

4. *Cadenas franquiciadas*. Es una forma de exportación directa bien adaptada a la distribución de determinados productos. La ventaja es la rápida expansión con débiles inversiones por parte del promotor.

A la hora de desarrollar la empresa de distribución comercial una estrategia de internacionalización pueden seguir dos enfoques diferentes: un enfoque global o un enfoque multinacional.

1. *Enfoque global*. Supone la reproducción de una fórmula en el extranjero (como ocurre generalmente con las cadenas franquiciadas). El objetivo reside en segmentos de consumidores que participan de unos gustos y estilos de vida similares con independencia de su país de origen. La adoptan empresas claramente centralizadas que procuran aprovechar economías de escala a través de unos modernos sistemas logísticos y de información.
2. *Enfoque multinacional*. Según la definición de Santesmases en [SM99], consiste en tratar de adaptar los puntos de venta a las condiciones nacionales de cada país donde se desea asentar. Tiene como ventaja principal su mayor efectividad, debido a que el ajuste a las características y necesidades de los mercados a los que se dirige es mucho mayor, pero como desventaja se debe resaltar su coste, que es más elevado.

El fenómeno de la internacionalización de empresas de distribución comercial es relativamente frecuente, sobre todo si se comparara a la internacionalización de empresas productoras de bienes de consumo. Este fenómeno se está viendo reforzado por el hecho de que determinadas formas comerciales se encuentran en fase de saturación y declive en sus países de origen y mayor desarrollo, como es el caso de las grandes tiendas en Chile que se encuentra en el proceso de mayor desarrollo, estimulándose a estas empresas a la

exportación de estos formatos hacia países en los que aún existe un potencial importante de crecimiento.

La internacionalización se puede combinar simultáneamente con otras estrategias de crecimiento. Además de la estrategia de penetración del mercado, se pueden aplicar estrategias de integración vertical y estrategias de diversificación.

D3.8. Estrategia de integración vertical

Tal como lo explica Santesmases en [SM99], el sistema de distribución tradicional comprende un fabricante independiente, un mayorista y un minorista. Cada uno de ellos pertenece a una entidad de negocio separada que busca maximizar sus propios beneficios, incluso aunque ello reduzca el del sistema en su totalidad. Ningún miembro del canal tiene un control completo o sustancial sobre los demás miembros. Ante este problema, los miembros del canal pueden decidir integrarse para conseguir de forma más eficiente la necesaria coordinación entre los objetivos de los intermediarios que forman el canal. En este sentido, la integración vertical tiene lugar cuando la función correspondiente a dos o más miembros del canal, situados a distinto nivel, es dirigida o administrada por alguno de sus componentes. El resultado de la integración vertical es una red de establecimientos, situados a distinto nivel del canal y dirigidos como un sistema de distribución centralizado, estrategia que opera como marco de referencia de las empresas bajo estudio.

Mediante la organización vertical de los canales de distribución, como lo explica Casares et al en [CJ96a], los participantes en las organizaciones buscan conseguir tanto la consecución de un mayor poder de mercado (hacia la demanda), así como un mayor poder de negociación (hacia los proveedores), por lo que los objetivos perseguidos por las empresas integradas en sistemas verticales podrían concretarse en los siguientes:

1. *Reducir la incertidumbre con la que operan las empresas, tanto en cuanto a aprovisionamientos (volatilidad de la oferta, alto poder de los oferentes) como en el lado de la demanda (poder creciente de los consumidores, escasa fidelidad de la clientela).*

2. *Lograr una mejor orientación de las inversiones* hacia actividades en el canal que presentan una mayor rentabilidad.
3. *La consecución de economías de escala y de alcance* por la coordinación de las funciones de distribución (optimización de los almacenamientos, y de las entregas de las mercancías, mejora del ciclo de pedidos y cobros, etc.).
4. *La creación de barreras de entrada*, pues los competidores potenciales verán aumentar la escala de operación a la vez que pueden encontrar dificultades para su aprovisionamiento, en función del poder alcanzado por las organizaciones verticales existentes en el mercado.

Por su misma definición, según Ansoff en [AI76], la diversificación es una estrategia arriesgada puesto que incluye una salida simultánea de productos y mercados familiares, frente a la estrategia de expansión de penetración en el mercado (crecimiento del negocio en el mismo producto-mercado), de desarrollo del mercado (crecimiento en base de búsqueda de nuevos mercados con el mismo producto) o de desarrollo del producto (crecimiento a través de ofrecer nuevos productos al mismo mercado). Dado que la decisión de diversificar o no, representa un hito importante en el desarrollo de la empresa será útil el estudiar, de forma general, las razones por las cuáles diversifican las empresas:

1. *Las empresas diversifican cuando no pueden alcanzar sus objetivos dentro del ámbito producto-mercado por la expansión.* En el área de los objetivos de rentabilidad a corto y a largo plazo la causa puede ser la saturación del mercado, el declive general de la demanda, las presiones competitivas o la obsolescencia de la línea de productos. En el área del objetivo flexibilidad la causa puede ser una parte desproporcionadamente grande de las ventas a un solo cliente, un mercado o base tecnológica generalmente reducido, o la influencia de las nuevas tecnologías en el campo producto-mercado de la empresa.

2. *Una empresa puede diversificar porque el disponible retenido excede las necesidades totales de expansión.* La rentabilidad que se puede obtener de los recursos líquidos (bancos, acciones) generalmente es más bajo que el de operaciones. Puede haber presiones para que la empresa invierta el dinero de una manera más rentable.
3. *Las empresas diversifican cuando las oportunidades de diversificación prometen mayor rentabilidad que las oportunidades expansivas.* Esto puede ocurrir cuando las oportunidades de diversificación son lo suficientemente atractivas para compensar su inherente menor sinergia, o cuando la sinergia no se considera importante y por tanto las ventajas de expansión de la sinergia sobre las de la diversificación no son significativas.
4. *Las empresas podrán diversificar cuando la información disponible no es lo suficientemente fiable para hacer una comparación concluyente entre expansión y diversificación.* Esto sucede muy a menudo, ya que las empresas tienen mucha más información respecto a proyectos de expansión que respecto al vasto campo externo de la diversificación.

Desde el punto de vista de las empresas de distribución comercial, la estrategia de diversificación comercial consiste en la entrada de una empresa distribuidora en nuevas líneas de actividad. Esta estrategia se puede llevar a cabo bien mediante desarrollo interno, o bien mediante adquisiciones. Según Muñiz en [MN95], las razones que mueven a una empresa de distribución comercial a iniciar una estrategia de diversificación pueden concretarse en las tres siguientes:

1. *Deseo de aprovechar los recursos y habilidades de la empresa* (activos intangibles como conocimientos del personal de la empresa; activos tangibles como redes de distribución; sinergias en la comunicación; aprovisionamiento más voluminoso que otorga un mayor poder de negociación; imagen de marca de la empresa, etc.).

2. *Búsqueda de nuevos negocios ante la mala perspectiva de los actuales.* Esta es una de las principales motivaciones de la búsqueda de crecimiento en las empresas de distribución detallista.

3. *Requerimientos legales que limitan la expansión de determinadas fórmulas comerciales.*

La diversificación comercial de una empresa de distribución puede tener diferentes niveles en función a la proximidad o alejamiento de su negocio básico. En este sentido, se pueden distinguir tres niveles de actuación:

1. *Diversificación en otros formatos comerciales relacionados con el negocio básico.*

2. *Diversificación en otros sectores comerciales.*

3. *Diversificación en sectores no relacionados (no comerciales).*

El primer nivel consiste en la diversificación en otros formatos comerciales relacionados con el negocio básico, teniendo como objetivo cubrir necesidades de consumo distintas de las que representa su fórmula originaria. En esta situación el riesgo inherente al cambio de negocio distinto del básico se minimiza, ya que se aprovechan los recursos y habilidades de la empresa, debido a realizarse la diversificación en ámbitos de gran proximidad al negocio básico. En este sentido, las principales empresas de tiendas por departamento asentadas en Chile están introduciendo un nuevo formato en sectores comerciales consolidados con tiendas que privilegian las marcas propias de las compañías.

El segundo nivel consiste en la diversificación hacia otro tipo de fórmulas comerciales con un dominio distinto al negocio básico. Dicho movimiento está encaminado a cubrir necesidades de consumo comercial totalmente distintas, como por ejemplo, la introducción en el sector Mejoramiento del Hogar, (Falabella, CencoSud), Farmacias

(Falabella, Ripley), y combustibles con el comercio asociado, etc. Debido al alejamiento del negocio básico, la diversificación en otros sectores comerciales diferentes al negocio básico supone asumir un riesgo más elevado por parte de las empresas.

Finalmente, el tercer nivel consiste en entrar en sectores no relacionados con la actividad comercial. En este caso, el riesgo puede ser muy elevado al no poder convalidarse siempre las habilidades y el saber hacer de la empresa distribuidora en este nuevo ámbito. Las empresas venta por departamento que operan en Chile han utilizado esta última estrategia de forma limitada. La diversificación menos relacionada con el negocio básico en las empresas de distribución comercial de vestuario y electro-domestico que operan en Chile se encamina principalmente hacia actividades complementarias a la puramente comercial o hacia actividades de apoyo. En este sentido, en el sector financiero, con la incursión en la Banca como lo realizado por Falabella, Ripley y Paris; en el sector inmobiliario comprando parte de la propiedad de los recintos comerciales que ocupan las tiendas (Centros Comerciales o Mall) y también podría diversificarse hacia actividades complementarias, como agencias de viajes y bancos.

D4. La atractividad de la industria del retail y las grandes tiendas

El análisis del sector industrial y de las actividad comercial de distribución de productos de vestuario y hogar en el formato de grandes tiendas deja en claro la creciente competitividad en la industria del retail, la que continúa e incrementa su agresividad se percibe en los recientes acontecimientos de este sector registrados en los artículos de prensa citados en [EM04a, 04b, 04c,04d,06], [LT04, 05] y [DF03], siendo uno de los principales negocios en este ámbito la venta a crédito y una de las principales estrategias comerciales la emisión por cuenta propia a sus clientes de tarjetas de créditos, tal como se aprecia en la publicación periodística [EM04a], la que señala que tres grandes tiendas lideran la industria del retail en el cono sur de América Latina (Argentina, Chile, Perú). En Chile ellas se disputan el 95% de las ventas, las que en 2005, según estas mismas publicaciones, superan los tres mil trescientos millones de dólares y sus participaciones de mercado en 2003 fueron 60,29%; 18,26% y 15,63% para las tres primeras, mientras que en 2004, estas cadenas totalizaron ventas cercanas a US\$2.600 millones, con

participaciones de mercado de 36,6%, 25,6%, 28,0% y 9,8%, para las cuatro primeras. En el área financiera, con alrededor de 10 millones de tarjetas de crédito emitidas y 60% de ellas activas, las cuatro tiendas por departamento alcanzaron colocaciones brutas por \$843 mil millones a fines de 2004, cifra que representa un 13% de crecimiento con respecto al 2003, mientras que en 2005, superó los US\$3.000 según las publicaciones especializadas [JR05] y [FR03, 04,05].

Según estas empresas especializadas, en 2002, las principales tiendas habían emitido 3,0; 2,7; 2,2; y 1,2 millones de tarjetas de crédito, sólo en Chile existen un total de 16 millones de tarjetas de crédito emitidas por las diferentes cadenas comerciales del retail, este tipo de tarjetas de crédito supera en 7 veces el número de tarjetas de crédito emitidas por las instituciones bancarias, el uso de ellas, en promedio alcanza al 65% de las ventas de sus casas emisoras, en total representa casi el 20% de la deuda de consumo de Chile, existiendo más de 11.000 establecimientos, no emisores de este tipo de tarjetas que operan con ellas.

Si bien es cierto la emisión masiva de tarjetas de crédito por parte de las grandes tiendas ha sido exitosa como proyecto de marketing, no es menos cierto, que junto con ello aumenta la exposición al riesgo de actividad ilícita, dada la tendencia que muestra la actividad fraudulenta como lo señalan las publicaciones especializadas como por ejemplo el último reporte de Cybersource [CS06]. La diversificación de la cartera de clientes con esta emisión masiva de tarjetas de crédito y los agresivos planes de marketing que incentivan el uso diverso de este medio de pago sumado a la carencia de técnicas eficientes y sistemas inteligentes que permitan detectar y prevenir en forma eficaz su uso ilegal, sin perjudicar naturalmente a quienes hagan uso genuino de este medio de pago conllevan al desafío de buscar métodos más eficientes, este esfuerzo se ve reflejado en distintos trabajos[BP03] y en particular los relacionados con el tema bajo estudio en las publicaciones especializadas, que aportan diferente enfoques para detectar y prevenir este comportamiento ilícito, no obstante, todas ellas se suman a lo indicado en el año 2002 por Bhatla[BP03], que indica que los sistemas evaluados adolecen de efectividad garantizada, y señala que ninguna de las herramientas y tecnologías revisadas puede por sí misma eliminar el fraude, y que cada técnica agrega valor a la habilidad de detectarlo, y postula que una buena práctica ha sido la combinación de varias de ellas, lo que aparentemente se mantiene como tendencia, dado el resultado de la encuesta hecha por Cybersource[CS06], en donde se señala que el control manual es el más utilizado en la detección y prevención del fraude, confirmando que la tendencia no ha cambiado.

E. Procedimiento para determinar el Perfil de la empresa, su Posicionamiento de Mercado y el Modelo de negocios adoptado.

A continuación se presenta el procedimiento para determinar el Perfil Comercial de la empresa y el Modelo de negocios que se ha dado para insertarse en la industria y en el medio ambiente donde se desenvuelve.

Para determinar el Perfil de la Empresa y su Modelo de Negocios, se propone una metodología que va de lo general a lo particular (TOP-DOWN).

La metodología que se propone tiene relación con determinar el posicionamiento de la empresa en su mercado y su nivel de participación, conocer y entender como funciona su organización de cara a sus clientes, la forma que adopta para generar sus ingresos, conocer como organiza su cadena de valor y como enfrenta las oportunidades y las amenazas teniendo en cuenta sus debilidades y fortalezas.

Ello significa analizar desde el macro entorno hacia la empresa, entendiendo que la empresa es un subsistema inserto en el Sistema Industria (el Sistema Industria lo integran todas aquellas empresas que resuelven la misma necesidad para un segmento objetivo determinado) y que a su vez el Sistema Industria es un subsistema del suprasistema denominado Macroentorno, tal como, se grafica en la figura E.1, siguiente:

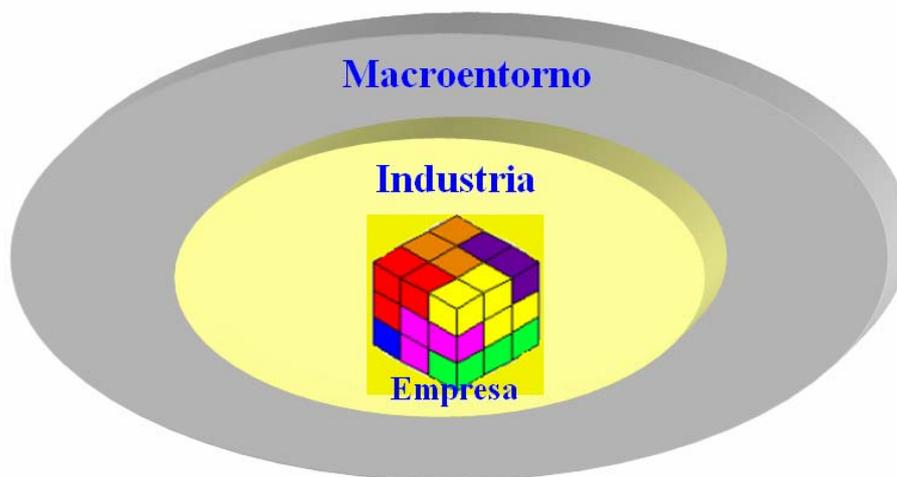


Fig. E.1 Sistema: Macroentorno, Industria, Empresa.

Para realizar este análisis se recomienda realizar dos Análisis, uno externo y otro interno a la empresa. Para realizar el escrutio externo se recomienda utilizar los siguientes dos modelos:

En primer lugar el Modelo PEST propuesto Andrews [AK84], que recomienda realizar un Análisis al *conjunto de condiciones e influencias externas que afectan o podrían afectar la vida y el desarrollo de una empresa como son los Factores PEST* que corresponden a las iniciales de los factores Políticos, Económicos, Sociales y Tecnológicos.

En segundo lugar, se propone realizar el análisis de la atractividad de la industria aplicando el modelo de análisis de las cinco fuerzas de Porter [PM92], modelo que analiza la situación actual y una estimación de la situación futura de cómo la empresa es afectada por las cinco fuerzas de la industria, estas son: Los Clientes; Los Proveedores; Los Sustitutos; Los competidores; y los Nuevos Entrantes.

Para realizar el Análisis interno de la Empresa se propone usar lo siguiente:

En primer lugar la matriz FODA, acrónimo para Fortaleza, Oportunidades, Debilidades, Amenazas, también denomina DOFA o también según sus siglas en inglés SWOT o DAFO, fue desarrollado por Marion Doshier, Dr. Otis Benepe, Robert Stewart y Birger Lie, siendo Albert S. Humphrey especialista en administración organizacional su principal exponente; las investigaciones fueron desarrolladas por el Instituto de Investigaciones de Standford (Standford Research Institute) en la década de los '60 a los '70. Financiada por las empresas del Fortune 500 para crear un sistema de análisis administrativo superior a la planificación corporativa de ese tiempo.

La segunda componente propuesta es el modelo Cadena de Valor de la empresa, que en su forma más tradicional Michael Porter [PM92] la organiza en dos grupos: Actividades de Apoyo y Actividades Primarias.

La tercera y última componente que se recomienda es el Modelo Delta, que permite comprender la visión, la misión y la estrategia de posicionamiento de la empresa, entre las cuales están: el mejor producto mediante las estrategias de diferenciación o de liderazgo en costo; la solución integral al cliente mediante la Amplitud horizontal o la redefinición de la experiencia con el cliente o la integración con el cliente; consolidación del sistema debiendo elegir entre Mercado dominante, Canal exclusivo o Estándar de Propiedad Registrado.

E1. Proceso para Determinar el Perfil de la Empresa

El proceso para determinar el Perfil de la empresa se inicia con el Análisis Externo que cubre el Análisis del Macro Entorno aplicando el

modelo PEST y se concluye con el Análisis de la atraktividad de la Industria.

Análisis Externo:

El análisis PEST permite conocer como las macro variables afectan a la empresa, determinando que tan favorable es el estado actual y futuro, esto es, favorable, neutro o desfavorable para la empresa dependiendo del estado y de la tendencia de cuatro macrovariables: la económica; la social; la política; y la tecnológica.

Cada una de ellas, afecta de manera diferente a los distintos tipos de organizaciones y estas condiciones habitualmente no son constantes, por lo que los escenarios que se generan para cada empresa se elaboraran para el presente y para el mediano plazo.

En la perspectiva económica, por ejemplo, una paridad cambiaría baja y tendiente a la baja afectará positivamente en el presente y en el futuro inmediato al importador y de manera negativa al exportador,

En cuanto a la variable política, que se refiere a la normativa vigente y en trámite, afectará de diferente manera a la organización, por ejemplo un gobierno proclive a la protección del medio ambiente será menos permisivo con las empresas productivas contaminantes, situación que puede ser permanente en el tiempo.

La variable social que tiene relación con las condiciones generales de la población, si existe un clima social que aprecie la contribución de la organización a la sociedad y sus prioridades, será favorable para el desarrollo de la organización, por ejemplo, una sociedad que considera prioritaria la educación, o la nutrición será más favorable para una empresa cuyo objetivo sean la prestación de servicios educacionales o la producción de alimentos con ciertas características nutricionales positivas, en este contexto social, también incide la componente demográfica, que permite conocer y determinar que tan favorable es el comportamiento de las poblaciones, por ejemplo el aumento de la población anciana o la disminución de la población infantil y el impacto positivo o negativo que ello puede tener para el mercado objetivo de la empresa en la actualidad y en el mediano y largo plazo.

La variable tecnológica permite determinar como será afectada la empresa por la tecnología emergente o la obsolescencia tecnológica, esto puede afectar en forma positiva o negativa tanto a la industria como a la empresa en particular.

Los resultados del análisis PEST permiten conocer y comprender cuales son los criterios provenientes del macro entorno que los expertos o los especialista consideran pertinentes al momento de tomar decisiones y como estas pueden afectar a la industria y a la empresa, dada su interaccione constante y permanente que se deduce de la siguiente figura E.2.



Fig. E.2 Componentes de Sistema: Macroentorno, Industria, Empresa.

Una vez conocido y entendido como afecta el macro-entorno a la empresa, tanto para el presente como para el futuro. Al estimar el estado actual y las tendencias de las variables relacionadas con la economía, la política, la sociedad y la tecnología, se inicia el análisis de la atractividad de la industria, entendiendo por industria al conjunto de empresas que producen el mismo satisfactor.

Para este proceso se utiliza el modelo de análisis de las cinco fuerzas de Porter [PM92], modelo que considera, al igual que el modelo de la lógica de escenario, un análisis de la situación actual y una estimación de la situación futura de cómo la empresa es afectada por las cinco fuerzas de la industria, estas son: Los Clientes; Los Proveedores; Los Sustitutos; Los competidores; y los Nuevos Entrantes.

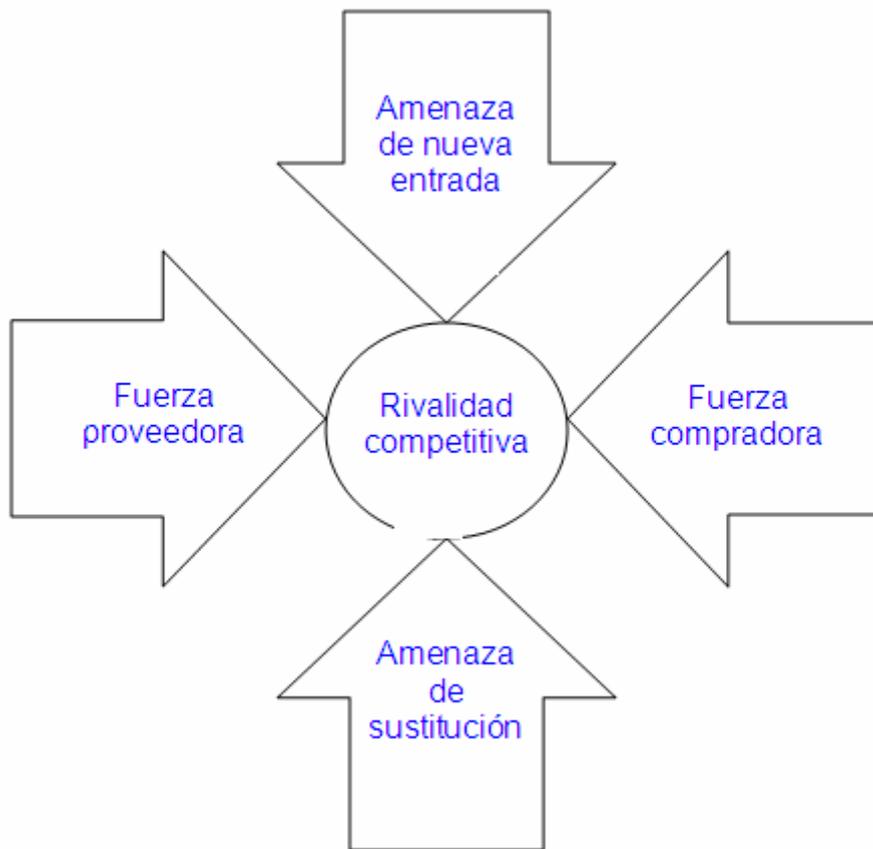


Fig. E.3 El modelo de las 5 fuerzas de la Industria de Porter

Para cada una de estas cinco fuerzas de la industria, se evalúa como afectan o afectara en el futuro a la organización, pudiendo ser éstas favorables, neutras o desfavorables tanto para el presente como para el futuro frente a las oportunidades y las amenazas que se han determinado en el Análisis externo.

Como resultados del Análisis externo se debería reconocer las oportunidades y las amenazas que genera el macro entorno y la industria en aspectos como los que se señalan a continuación:

Oportunidades

- ¿Efectos políticos?
- ¿Efectos legislativos?
- ¿Efectos ambientales?
- ¿Desarrollos de TI?
- Desarrollos del mercado?
- ¿Vulnerabilidades de los competidores?
- ¿Tendencias de la industria o de estilo de vida?
- ¿Desarrollos tecnológicos e innovaciones?
- ¿Influencias globales?
- ¿Nuevos mercados, verticales, horizontales?
- ¿Mercados objetivo nicho?
- ¿Geografía, exportación, importación?
- ¿Nuevas propuestas únicas de venta?

- ¿Tácticas - sorpresa, grandes contratos, etc?
- ¿Desarrollo de negocios o de productos?
- ¿Información e investigación?
- ¿Sociedades, agencias, distribución?
- ¿Volúmenes, producción, economías?
- ¿Influencias estacionales, del clima, o de la moda?

Amenazas

- ¿Efectos políticos?
- ¿Efectos legislativos?
- ¿Efectos ambientales?
- ¿Desarrollos de TI?
- ¿Intenciones de los competidores?
- ¿Demanda del mercado?
- ¿Nuevas tecnologías, servicios, ideas?
- ¿Contratos y alianzas vitales?
- ¿Mantener las capacidades internas?
- ¿Obstáculos enfrentados?
- ¿Debilidades no superables?
- ¿Pérdida de personal clave?
- ¿Respaldo financiero sostenible?
- ¿Economía – local o extranjera?
- ¿Influencias estacionales, del clima, o de la moda?

E2. Proceso para Determinar el Modelo de Negocios

Análisis Interno:

Una vez concluido el proceso anterior, denominado Análisis externo, se procede a realizar el denominado Análisis interno, con el propósito de verificar donde la empresa presenta fortalezas y debilidades, para ello se recomienda conocer y entender la cadena de valor de la organización,

La cadena de valor es la forma en que cada empresa organiza sus procesos, es la forma en la cual la empresa pretende crear valor y generar diferenciación con la competencia para sus clientes, es por ello, la importancia de conocer y entender el funcionamiento de la cadena de valor de la empresa, que en su forma más tradicional Michael Porter la gráfica de la forma que la muestra la figura E.4, siguiente:

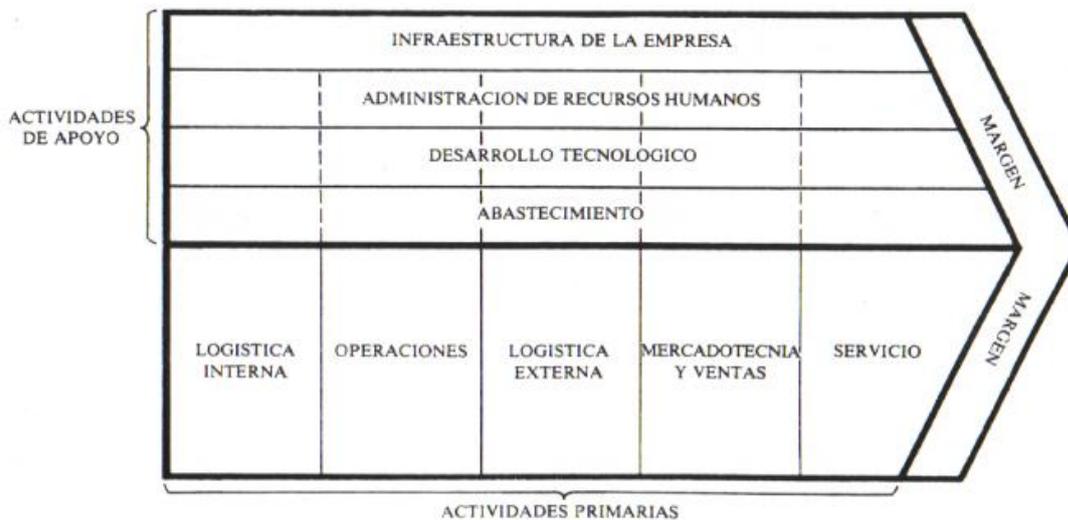


Fig. E.4 Cadena de Valor. Fuente: Michael Porter[PM92].

En la gráfica anterior se aprecia como se distribuyen las actividades de la empresa, las que se organizan básicamente en dos grupos:

Las actividades de apoyo: que corresponde a todas aquellas actividades que contribuyen a la generación del producto o servicio, muchas de ellas son susceptibles de externalizar o encargar a terceros que puedan hacer mejor estas actividades que la propia empresa, ya sea, por ser especialistas en el rubro o por tener la posibilidad de producir economías de escala, como por ejemplo en la informática, arrendar espacio y servicios de datacenter, en lugar de tener que construir y administrar uno propio, o los generales, tales como: servicios de vigilancia o aseo, servicios especializados como la selección, reclutamiento y administración de personal.

Las actividades primarias: que corresponden a todas aquellas actividades propias del giro de la empresa y relacionadas con la generación del producto o la prestación del servicio de una forma que caracteriza a la empresa y le otorga un sello distintivo frente a sus clientes y una diferenciación frente a sus competidores.

Para concluir este Análisis interno, se debe conocer las definiciones estratégicas de la organización se ha dado en términos de producto, mercado objetivo y tecnología, entendiendo por producto, al resultado que la empresa ofrece al cliente, sea este un bien o un servicio con el valor agregado o sello de la empresa, por mercado objetivo al conjunto de clientes al cual esta dirigido el producto y la tecnología a la forma adoptada por la empresa para producir su producto o servicios.

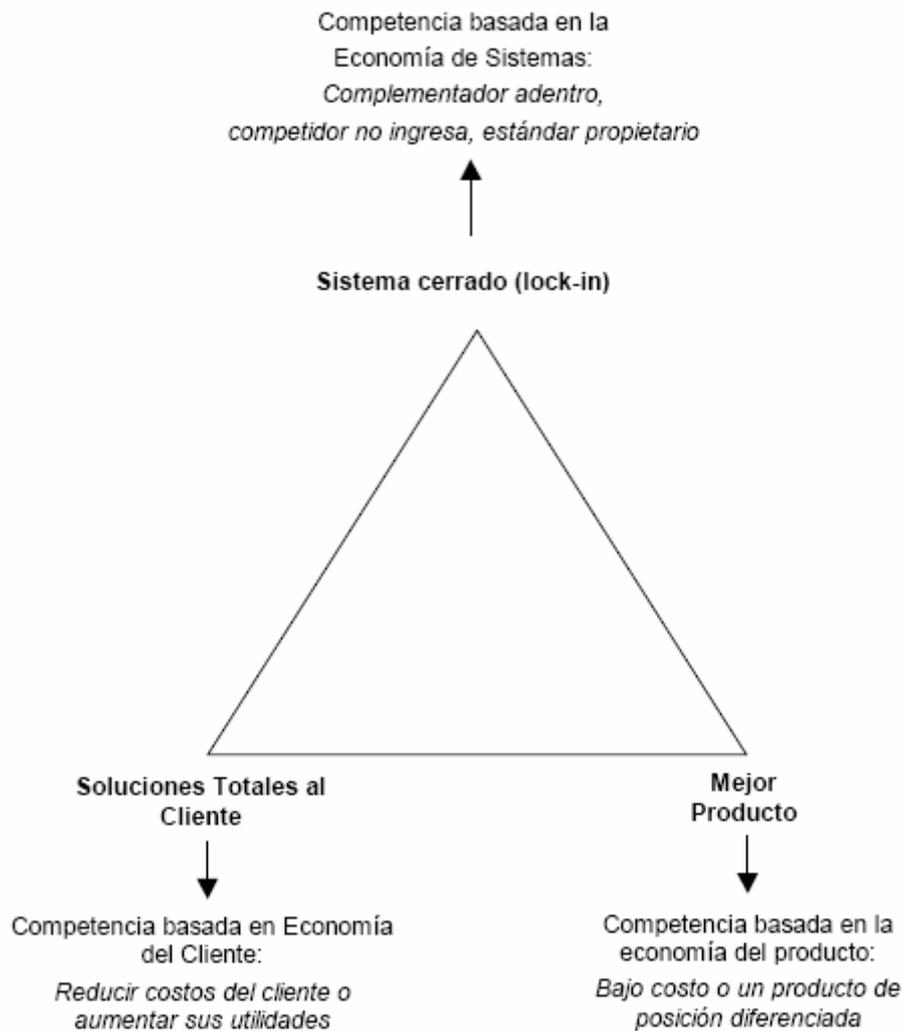


Fig. E. 5 Modelo Comercial: Tres Perspectivas Estratégicas. Fuente: Arnoldo Hax[HA01].

Estas definiciones, por lo habitual, se encuentran en la declaración de misión, visión, objetivos estratégicos, metas de la organización y las estrategias elegidas para su cumplimiento.

Habitualmente estas definiciones son consecuencia de un posicionamiento estratégico que corresponden a alguno de los vértices descritos por el Modelo Delta de Hax[HA01], que se muestran y explica la figura E.5.

Como resultados del Análisis interno se debería reconocer las fortalezas y las debilidades que se generan al interior de la empresa en aspectos como los que se señalan a continuación:

Fortalezas

- ¿Ventajas frente a las amenazas y oportunidades?
- ¿Capacidades?
- ¿Ventajas competitivas?
- ¿PUV's (propuesta única de vetas)?
- ¿Recursos, activos, gente?

- ¿Experiencia, conocimiento, datos?
- ¿Reservas financieras, retorno probable?
- ¿Marketing – alcance, distribución, awareness?
- ¿Aspectos innovadores?
- ¿Ubicación geográfica?
- ¿Precio, valor, calidad?
- ¿Acreditaciones, calificaciones, certificaciones?
- ¿Procesos, sistemas, TI, comunicaciones?
- ¿Cultural, actitudinal, de comportamiento?
- ¿Cobertura gerencial, sucesión?

Debilidades

- ¿Desventajas frente a las amenazas y oportunidades?
- ¿Brechas en la capacidad?
- ¿Falta de fuerza competitiva?
- ¿Reputación, presencia y alcance?
- ¿Aspectos Financieros?
- ¿Vulnerabilidades propias conocidas?
- ¿Escala de tiempo, fechas tope y presiones?
- ¿Flujo de caja, drenaje de efectivo?
- ¿Continuidad, robustez de la cadena de suministros?
- ¿Efectos sobre las actividades principales, distracción?
- ¿Confiabilidad de los datos, predictibilidad del plan?
- ¿Motivación, compromiso, liderazgo?
- ¿Acreditación, etc?
- ¿Procesos y sistemas, etc?
- ¿Cobertura gerencial, sucesión?

Estos procesos para conocer y estudiar el Perfil de la Empresa y el Modelo de Negocios, aplicando un procedimiento que va de lo general a lo particular, realizando Análisis internos y externos, permite además de conocer y entender la forma que la empresa se ha dado para insertarse en la industria, permite interiorizarse del glosario interno que emplean los analistas o los especialistas para explicar los conceptos, las variables y la pertinencia que éstos le otorgan a cada una de ellos en el proceso de toma de decisiones.

E3. Formularios para registrar el Análisis de cada nivel

Macroentorno						
	Condiciones Actuales			Condiciones Futuras		
	Desfavorable	Neutro	Favorable	Desfavorable	Neutro	Favorable
Político						
Social						
Económico						
Demográfico						
Escenario						

Fig. E. 6 Síntesis del Análisis del Macro Entorno

Industria						
	Condiciones Actuales			Condiciones Futuras		
	Desfavorable	Neutro	Favorable	Desfavorable	Neutro	Favorable
Competidores						
Sustitutos						
Proveedores						
Clientes						
Nuevos Entrantes						
Atractividad Industria						

Fig. E. 7 Síntesis del Análisis de la Industria

Empresa

	Condiciones Actuales			Condiciones Futuras		
	Desfavorable	Neutro	Favorable	Desfavorable	Neutro	Favorable
Organización						
Productos o Servicios						
Capital Intelectual						
Competitividad De la Empresa						

Fig. E. 8 Síntesis del Análisis de la Empresa

E4. Resumen del Procedimiento

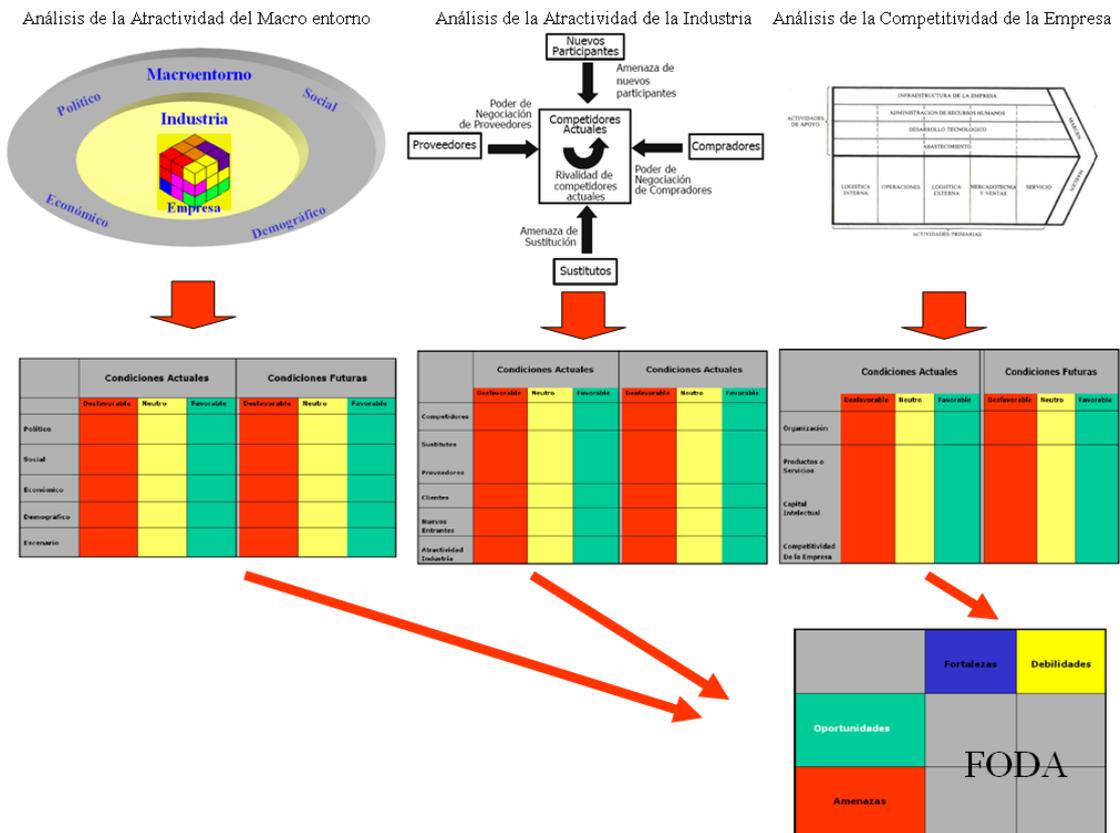


Fig. E.9 Procedimiento para determinar el Perfil de la empresa y su modelo de negocios

F. Resultados de la aplicación de la metodología FBI al caso de Gestión de Riesgo y Control Interno.

A continuación se presentan el detalle del resultado de la aplicación de la metodología propuesta al caso de Gestión de Riesgo y Control Interno.

La lista de ítems frecuentes de la muestra de datos es la siguiente:

Variable	F. Soporte
[TX_NEGOCIO_F = 'MEDIO']	100,00%
[TX_ENDISPUTA = 0]	99,11%
[CL_LCD_F = 'BAJO']	97,14%
[TX_M_LIQ_F = 'BAJO']	90,77%
[TX_CANT = 1]	90,00%
[TX_TIPO = 21]	80,21%
[TX_PLAZOF = 'CORTO']	67,71%
[TX_LPROD_F = 'MEDIO']	65,39%
[SEXO = 'F']	59,83%
[TX_TITUL = 1]	59,27%
[CAT_CL = 0]	58,32%
[CAT_CL = 1]	41,68%
[CL_NCD = 1]	41,08%
[SEXO = 'M']	40,17%
[CL_EDADF = 'ADULTO']	32,58%
[TX_CUOTAS = 2]	29,99%
[CL_EDADF = 'ADULTO JOVEN']	27,35%
[CL_EDADF = 'JOVEN']	24,40%
[TX_TITUL = 0]	23,12%
[TX_PLAZOF = 'ALTO']	21,79%
[CL_NCD = 2]	21,75%
[TX_CUOTAS = 1]	19,83%
[TX_LPROD_F = 'BAJO']	19,65%
[TX_TITUL = 2]	17,61%
[TX_CUOTAS = 3]	16,91%
[CL_EDADF = 'ADULTO MAYOR']	15,68%
[TX_LPROD_F = 'ALTO']	14,96%
[CL_NCD = 3]	11,96%
[TX_PLAZOF = 'MEDIANO']	10,50%
[TX_CUOTAS = 0]	10,50%
[CL_NCD = 4]	7,33%
[TX_M_LIQ_F = 'MEDIO']	6,97%
[TX_CUOTAS = 6]	6,90%
[TX_CUOTAS = 4]	6,15%

[TX_TIPO = 91]	5,59%
[CL_NCD = 5]	4,81%
[TX_CANT = 0]	4,76%
[TX_TIPO = 57]	4,27%
[TX_TIPO = 13]	3,82%
[CL_NCD = 6]	3,17%
[TX_CUOTAS = 5]	3,07%
[TX_CANT = -1]	2,92%
[CL_LCD_F = 'MEDIO']	2,85%
[TX_TIPO = 36]	2,76%
[CL_NCD = 7]	2,30%
[TX_M_LIQ_F = 'ALTO']	2,25%
[TX_CUOTAS = 12]	1,83%
[TX_TIPO = 32]	1,83%
[TX_CUOTAS = 10]	1,68%
[TX_CUOTAS = 8]	1,53%
[CL_NCD = 8]	1,45%
[CL_FTX = '21-12-2002']	1,26%
[TX_CANT = 2]	1,19%
[CL_FTX = '23-12-2002']	1,17%
[CL_NCD = 9]	1,12%
[CL_FTX = '24-12-2002']	1,07%

Tabla F.1. Items Frecuentes de Impugnaciones

Ejemplos del resultado de aplicar las medidas de interés objetivas para obtener reglas muy fuertes que cumplen la condición de $SuppA+SuppB > 1$:

A	->	C	Supp	Conf	FC	SuppA	SuppC	SuppA+C
CAT_CL = 0	->	TX_NEGOCIO_F = 'MEDIO',TX_ENDISPUTA = 0	58,322	100,000	1,000	0,583	0,991	1,574
CAT_CL = 0	->	TX_ENDISPUTA = 0,	58,322	100,000	1,000	0,583	0,991	1,574
CAT_CL = 0,TX_LPROD_F = 'MEDIO'	->	TX_NEGOCIO_F = 'MEDIO',TX_ENDISPUTA = 0	38,523	100,000	1,000	0,583	0,991	1,574
CAT_CL = 0,TX_LPROD_F = 'MEDIO'	->	TX_ENDISPUTA = 0,	38,523	100,000	1,000	0,583	0,991	1,574
CAT_CL = 0,TX_LPROD_F = 'MEDIO',TX_NEGOCIO_F = 'MEDIO'	->	TX_ENDISPUTA = 0	38,523	100,000	1,000	0,583	0,991	1,574
CAT_CL = 0,TX_NEGOCIO_F = 'MEDIO'	->	TX_ENDISPUTA = 0,	58,322	100,000	1,000	0,583	0,991	1,574
CL_EDADF = 'ADULTO',CL_LCD = 1	->	CL_LCD_F = 'BAJO',	31,552	100,000	1,000	0,323	0,971	1,295
CL_EDADF = 'ADULTO',CL_LCD = 1,TX_ENDISPUTA = 0	->	CL_LCD_F = 'BAJO'	31,321	100,000	1,000	0,323	0,971	1,295

CL_EDADF = 'ADULTO',CL_LCD = 1, TX_NEGOCIO_F = 'MEDIO'	->	CL_LCD_F = 'BAJO'	31,552	100,000	1,000	0,323	0,971	1,295
CL_EDADF = 'ADULTO',CL_LCD_F = 'BAJO'	->	CL_LCD = 1,	31,552	100,000	1,000	0,323	0,971	1,295
CL_EDADF = 'ADULTO',CL_LCD_F = 'BAJO',TX_ENDISPUTA = 0	->	CL_LCD = 1	31,321	100,000	1,000	0,323	0,971	1,295
CL_EDADF = 'ADULTO',CL_LCD_F = 'BAJO',TX_NEGOCIO_F = 'MEDIO'	->	CL_LCD = 1	31,552	100,000	1,000	0,323	0,971	1,295
CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_ENDISPUTA = 0	96,283	99,116	0,762	0,971	0,571	1,542
CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_NEGOCIO_F = 'MEDIO',TX_ENDISPUTA = 0	96,283	99,116	0,762	0,971	0,571	1,542
CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_NEGOCIO_F = 'MEDIO'	97,142	100,000	1,000	0,971	0,571	1,542
CL_LCD = 1	->	CL_LCD_F = 'BAJO',	97,142	100,000	1,000	0,971	0,971	1,943
CL_LCD = 1,CAT_CL = 0	->	CL_LCD_F = 'BAJO',TX_ENDISPUTA = 0	57,000	100,000	1,000	0,571	0,571	1,142
CL_LCD = 1,CAT_CL = 0	->	TX_NEGOCIO_F = 'MEDIO',TX_ENDISPUTA = 0	57,000	100,000	1,000	0,571	0,991	1,562
CL_LCD = 1,CAT_CL = 0	->	CL_LCD_F = 'BAJO',TX_NEGOCIO_F = 'MEDIO'	57,000	100,000	1,000	0,571	0,571	1,142

Tabla F. 2 Reglas Muy Fuertes que cumplen las condición SuppA+ SuppC > 1

Ejemplos del resultado de aplicar las medidas de interés objetivas para obtener reglas muy fuertes que cumplen la condición de $A+C < 1$:

A	->	C	Supp	Conf	FC	SuppA	SuppC	SuppA +C
CL_EDADF = 'ADULTO',CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_ENDISPUTA = 0	31,321	99,267	0,803	0,323	0,571	0,894
CL_EDADF = 'ADULTO',CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_NEGOCIO_F = 'MEDIO'	31,552	100,000	1,000	0,323	0,571	0,894
CL_EDADF = 'ADULTO',CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_ENDISPUTA = 0	31,321	99,267	0,803	0,323	0,571	0,894
CL_EDADF = 'ADULTO',CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_NEGOCIO_F = 'MEDIO'	31,552	100,000	1,000	0,323	0,571	0,894
CL_NCD = 1,CL_LCD = 1	->	CL_LCD_F = 'BAJO',TX_ENDISPUTA = 0	39,589	98,601	0,624	0,316	0,571	0,887

CL_NCD = 1, CL_LCD = 1	->	CL_LCD_F = 'BAJO', TX_NEGOCIO_F = 'MEDIO'	40,150	100,000	1,000	0,316	0,571	0,887
CL_NCD = 1, CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_ENDISPUTA = 0	39,589	98,601	0,624	0,316	0,571	0,887
CL_NCD = 1, CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_NEGOCIO_F = 'MEDIO'	40,150	100,000	1,000	0,316	0,571	0,887
CL_NCD = 1, TX_TIPO = 21	->	CL_LCD = 1, TX_CANT = 1	30,726	95,434	0,637	0,316	0,571	0,887
CL_NCD = 1, TX_TIPO = 21	->	CL_LCD_F = 'BAJO', TX_CANT = 1	30,726	95,434	0,637	0,316	0,571	0,887
SEXO = 'F', CL_LCD = 1	->	CL_LCD_F = 'BAJO', TX_ENDISPUTA = 0	57,826	99,150	0,771	0,413	0,571	0,984
SEXO = 'F', CL_LCD = 1	->	CL_LCD_F = 'BAJO', TX_NEGOCIO_F = 'MEDIO'	58,322	100,000	1,000	0,413	0,571	0,984
SEXO = 'F', CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_ENDISPUTA = 0	57,826	99,150	0,771	0,413	0,571	0,984
SEXO = 'F', CL_LCD_F = 'BAJO'	->	CL_LCD = 1, TX_NEGOCIO_F = 'MEDIO'	58,322	100,000	1,000	0,413	0,571	0,984
SEXO = 'F', TX_TIPO = 21	->	TX_TITUL = 1, TX_NEGOCIO_F = 'MEDIO'	39,993	84,106	0,610	0,413	0,586	1,000
SEXO = 'F', TX_TITUL = 1	->	TX_TIPO = 21, TX_CANT = 1	38,895	93,192	0,689	0,413	0,563	0,976

Tabla F. 3 Reglas Muy fuertes que cumplen la condición A+C < 1

Ejemplos de reglas muy fuertes que cumplen la condición $\neg C \Rightarrow \neg A$:

A	=>	C	Supp	Conf	FC	Supp A
CL_NCD = 2, TX_TIPO = 21	=>	TX_PLAZOF = 'CORTO'	17,915	99,450	0,983	0,180
TX_PLAZOF = 'ALTO'	=>	CL_NCD = 1	21,583	99,052	0,984	0,218
CL_NCD = 2, TX_TITUL = 1	=>	SEXO = 'F', TX_PLAZOF = 'CORTO'	9,598	70,940	0,511	0,135
SEXO = 'M', TX_PLAZOF = 'ALTO'	=>	CL_NCD = 1, CAT_CL = 0	5,468	61,927	0,502	0,088

Tabla F. 4 Reglas fuertes que cumplen la condición si $\neg C \Rightarrow \neg A$

Ejemplos de los ítemes frecuentes:

Variables	Soporte
[TX_NEGOCIO_F = 'MEDIO']	100,00%
[TX_ENDISPUTA = 1]	100,00%
[CAT_CL = 1]	100,00%
[TX_TIPO = 21]	99,07%
[TX_CANT = 1]	97,22%
[CL_LCD_F = 'BAJO']	96,30%

[TX_M_LIQ_F = 'BAJO']	77,78%
[TX_TITUL = 1]	72,22%
[CL_NCD = 1]	64,81%
[SEXO = 'F']	58,33%
[TX_LPROD_F = 'MEDIO']	52,78%
[TX_PLAZOF = 'CORTO']	48,15%
[SEXO = 'M']	41,67%
[TX_PLAZOF = 'ALTO']	35,19%
[CL_EDADF = 'ADULTO MAYOR']	31,48%
[TX_LPROD_F = 'BAJO']	28,70%
[CL_EDADF = 'ADULTO']	28,70%
[TX_CUOTAS = 3]	21,30%
[CL_EDADF = 'JOVEN']	21,30%
[TX_LPROD_F = 'ALTO']	18,52%
[CL_EDADF = 'ADULTO JOVEN']	18,52%
[TX_PLAZOF = 'MEDIANO']	16,67%
[TX_CUOTAS = 2]	15,74%
[TX_TITUL = 0]	14,81%
[TX_TITUL = 2]	12,96%
[TX_M_LIQ_F = 'ALTO']	12,04%
[TX_CUOTAS = 6]	12,04%
[CL_NCD = 2]	12,04%
[TX_CUOTAS = 1]	11,11%
[TX_M_LIQ_F = 'MEDIO']	10,19%
[TX_CUOTAS = 4]	10,19%
[TX_CUOTAS = 12]	8,33%
[CL_FTX = '21-12-2002']	7,41%
[TX_CUOTAS = 10]	6,48%
[CL_FTX = '28-06-2002']	6,48%
[CL_FTX = '23-12-2002']	6,48%
[TX_CUOTAS = 8]	5,56%

Tabla F. 5 Items Frecuentes

G. Resultados de la aplicación de la metodología FBI al caso de Gestión de Negocios.

A continuación se presentan el detalle del resultado de la aplicación de la metodología propuesta al caso de Gestión de Negocios.

La lista de ítems frecuentes de la muestra de datos es la siguiente:

Concepto	Etiqueta	Frecuencia
Stock Medio	SM	99,16
Kanban Sin variación	KS	99,13
Disponibilidad Cero	DC	93,18
Tienda Regular	TR	79,82
Venta Mala	VM	70,07
Item Cero Venta	IC	67,65
Kanban Cero	KC	62,22
Precio Rebajado	PR	60,71
Tienda en Local Grande	LG	57,6
Ciclo Pleno	CP	52,99
Tienda Ubicada en Mall	UM	41,66
Precio Lista	PL	39,29
Tienda Ubicada en RM	UC	39,02
Kanban Vigente	KV	37,78
Ciclo Final	CF	30,03
Venta Regular	VR	29,06
Tienda en Local Mediano	LM	23,75
Item Perdedor	IP	22,59
Tienda Ganadora	TG	20,18
Tienda Ubicada en Región	UR	19,32
Tienda en Local Pequeño	LP	18,66
Ciclo Inicial	CI	16,95
Item Ganador	IG	9,76
Disponibilidad Alta	DA	3,47
Disponibilidad Baja	DB	3,35
Venta Buena	VB	0,88
Stock Alto	SA	0,84
Disponibilidad Media	DM	0
Stock Bajo	SB	0
Item Quebrado	IQ	0
Item Baja Rotación	IB	0
Kanban Aumentado	KA	0
Kanban Disminuido	KD	0
Precio Aumentado	PA	0
Tienda Perdedora	TP	0
Venta Caída	VC	0

Tabla G. 1 Distribución de frecuencia según etiquetas lingüísticas

El conjunto de reglas de asociación obtenidas del proceso de extracción, se ordenan de mayor a menor por soporte, confianza y factor de certeza y se realiza un ejercicio de comparación en ese orden con el conocimiento aportado por el usuario, este proceso se aplica a las primeras 100 reglas con los siguientes resultados:

En primer lugar aparecen las reglas relacionadas con el Stock y la Venta y el conocimiento aportado por el usuario en este sentido indica:

“Si el stock no es alto la venta es mala”: $\neg SA \rightarrow VM$

Las siguientes reglas permiten validar la afirmación anterior:

A1	A2	A3	->	C1	C2	C3	supp	Conf	fcert
'VM'			->	'SM'			70,067116	100	1
'AM'			->	'SM'			69,99885	99,99104	0,9893181
'VM'	'AM'		->	'SM'			69,694435	100	1
'AM'	'SM'		->	'VM'			69,694435	99,56512	0,9854714
'AM'			->	'VM'	'SM'		69,694435	99,5562	0,98517334
'VM'			->	'AM'	'SM'		69,694435	99,46804	0,9822687
'VM'	'SM'		->	'AM'			69,694435	99,46804	0,98226506
'KS'	'AM'		->	'SM'			69,43333	100	1
'AM'			->	'KS'	'SM'		69,43333	99,1836	0,5186419
'KS'	'VM'		->	'SM'			69,42716	100	1
'KS'	'VM'	'AM'	->	'SM'			69,24702	100	1
'KS'	'VM'		->	'AM'	'SM'		69,24702	99,740555	0,9913521
'KS'	'VM'	'SM'	->	'AM'			69,24702	99,740555	0,9913503
'KS'	'AM'	'SM'	->	'VM'			69,24702	99,7315	0,9910299
'KS'	'AM'		->	'VM'	'SM'		69,24702	99,7315	0,9910299
'VM'	'AM'		->	'KS'	'SM'		69,24702	99,35831	0,62165123
'AM'	'SM'		->	'KS'	'VM'		69,24702	98,92601	0,96487105
'AM'			->	'KS'	'VM'	'SM'	69,24702	98,91734	0,96458757
'VM'	'SM'		->	'KS'	'AM'		69,24702	98,829636	0,961711
'VM'			->	'KS'	'AM'	'SM'	69,24702	98,829636	0,961711
'VM'	'DC'		->	'SM'			64,935326	100	1
'AM'	'DC'		->	'SM'			64,860886	99,990326	0,98846525
'VM'	'AM'	'DC'	->	'SM'			64,59997	100	1
'AM'	'SM'	'DC'	->	'VM'			64,59997	99,59774	0,9865613
'AM'	'DC'		->	'VM'	'SM'		64,59997	99,58807	0,9862383
'VM'	'DC'		->	'AM'	'SM'		64,59997	99,48344	0,9827819
'VM'	'SM'	'DC'	->	'AM'			64,59997	99,48344	0,9827783
'AM'	'SM'		->	'VM'	'DC'		64,59997	92,28724	0,78004175
'AM'			->	'VM'	'SM'	'DC'	64,59997	92,279045	0,7798082
'VM'			->	'AM'	'SM'	'DC'	64,59997	92,19721	0,7779457
'VM'	'SM'		->	'AM'	'DC'		64,59997	92,19721	0,77790684
'KS'	'VM'	'DC'	->	'SM'			64,34518	100	1
'KS'	'AM'	'DC'	->	'SM'			64,3265	100	1
'AM'	'DC'		->	'KS'	'SM'		64,3265	99,16682	0,5087455

Tabla G.2 Reglas relacionadas con el Stock y la venta

Luego se encuentran las reglas relacionadas con Disponibilidad del producto, en este sentido, el conocimiento del usuario señala:

“Si no hay disponible y el stock no es alto la venta es mala”: DC – SA
-> VM

Verificamos que las siguientes reglas validan la afirmación anterior:

A1	A2	A3	->	C1	C2	C3	Supp	conf	fcert
'VM'	'DC'		->	'SM'			64,935326	100	1
'AM'	'DC'		->	'SM'			64,860886	99,990326	0,98846525
'VM'	'AM'	'DC'	->	'SM'			64,59997	100	1
'AM'	'SM'	'DC'	->	'VM'			64,59997	99,59774	0,9865613
'AM'	'DC'		->	'VM'			64,59997	99,58807	0,9862383
'AM'	'DC'		->	'VM'	'SM'		64,59997	99,58807	0,9862383
'VM'	'DC'		->	'AM'	'SM'		64,59997	99,48344	0,9827819
'VM'	'DC'		->	'AM'			64,59997	99,48344	0,9827783
'VM'	'SM'	'DC'	->	'AM'			64,59997	99,48344	0,9827783
'AM'	'SM'		->	'VM'	'DC'		64,59997	92,28724	0,78004175
'AM'			->	'VM'	'DC'		64,59997	92,279045	0,7798082
'AM'			->	'VM'	'SM'	'DC'	64,59997	92,279045	0,7798082
'VM'			->	'AM'	'SM'	'DC'	64,59997	92,19721	0,7779457
'VM'			->	'AM'	'DC'		64,59997	92,19721	0,77790684
'VM'	'SM'		->	'AM'	'DC'		64,59997	92,19721	0,77790684
'KS'	'VM'	'DC'	->	'SM'			64,34518	100	1
'KS'	'AM'	'DC'	->	'SM'			64,3265	100	1
'AM'	'DC'		->	'KS'	'SM'		64,3265	99,16682	0,5087455
'KS'	'AM'	'DC'	->	'VM'			64,1774	99,76812	0,99225336
'KS'	'VM'	'DC'	->	'AM'			64,1774	99,739265	0,9913074
'AM'	'DC'		->	'KS'	'VM'		64,1774	98,93698	0,9652299
'VM'	'DC'		->	'KS'	'AM'		64,1774	98,83272	0,961812
'KS'	'VM'		->	'AM'	'DC'		64,1774	92,4385	0,7847748
'KS'	'AM'		->	'VM'	'DC'		64,1774	92,43014	0,78411716
'AM'			->	'KS'	'VM'	'DC'	64,1774	91,67556	0,766527
'VM'			->	'KS'	'AM'	'DC'	64,1774	91,594185	0,76436806

Tabla G.3. Reglas relacionada con la disponibilidad del producto

A continuación aparecen las reglas relacionadas con los ítems quebrados o ítems en cero, en este sentido, el conocimiento del usuario señala:

“Si el item esta quebrado la venta es mala”: IC -> VM

Verificamos que las siguientes reglas validan la afirmación anterior:

A1	A2	A3	->	C1	C2	C3	Supp	conf	Fcert
'IC'			->	'KC'			62,164593	91,890785	0,7853548
'KC'	'SM'		->	'IC'			61,82287	99,90973	0,9972095
'KC'			->	'IC'	'SM'		61,82287	99,360886	0,9804499
'IC'	'SM'		->	'KC'			61,82287	91,84972	0,7842676
'IC'			->	'KC'	'SM'		61,82287	91,38572	0,77402997

'KC'	'KS'		->	'IC'			61,80419	99,90973	0,9972095
'KC'			->	'KS'	'IC'		61,80419	99,33095	0,97958857
'KS'	'IC'		->	'KC'			61,80419	91,940865	0,7866802
'IC'			->	'KC'	'KS'		61,80419	91,3582	0,773418
'KC'	'KS'	'SM'	->	'IC'			61,462425	99,909164	0,997192
'KC'	'KS'		->	'IC'	'SM'		61,462425	99,35736	0,98034215
'KC'	'SM'		->	'KS'	'IC'		61,462425	99,32728	0,9794766
'KC'			->	'KS'	'IC'	'SM'	61,462425	98,781784	0,96321803
'KS'	'IC'	'SM'	->	'KC'			61,462425	91,89967	0,78559005
'KS'	'IC'		->	'KC'	'SM'		61,462425	91,43245	0,7752558
'IC'	'SM'		->	'KC'	'KS'		61,462425	91,314445	0,77227074
'IC'			->	'KC'	'KS'	'SM'	61,462425	90,85313	0,7623049
'KC'	'DC'		->	'IC'			58,50524	99,90459	0,9970505
'KC'			->	'IC'	'DC'		58,50524	94,02902	0,8359635
'IC'	'DC'		->	'KC'			58,50524	91,98988	0,7879777
'IC'			->	'KC'	'DC'		58,50524	86,48174	0,673779
'KC'	'SM'	'DC'	->	'IC'			58,188324	99,90392	0,9970301
'KC'	'DC'		->	'IC'	'SM'		58,188324	99,36351	0,9805303
'KC'	'SM'		->	'IC'	'DC'		58,188324	94,03611	0,83615816
'KC'			->	'IC'	'SM'	'DC'	58,188324	93,519646	0,8235064
'IC'	'SM'	'DC'	->	'KC'			58,188324	91,94968	0,7869136
'IC'	'DC'		->	'KC'	'SM'		58,188324	91,491684	0,77680945
'IC'	'SM'		->	'KC'	'DC'		58,188324	86,450134	0,6730162
'IC'			->	'KC'	'SM'	'DC'	58,188324	86,01331	0,66503507
'KC'	'KS'	'DC'	->	'IC'			58,176018	99,903915	0,9970299
'KC'	'DC'		->	'KS'	'IC'		58,176018	99,3423	0,9799348
'KC'	'KS'		->	'IC'	'DC'		58,176018	94,04444	0,8363871
'KC'			->	'KS'	'IC'	'DC'	58,176018	93,49975	0,82335323
'KS'	'IC'	'DC'	->	'KC'			58,176018	92,04763	0,7895064
'IC'	'DC'		->	'KC'	'KS'		58,176018	91,47208	0,7764037
'KS'	'IC'		->	'KC'	'DC'		58,176018	86,54334	0,67526555
'IC'			->	'KC'	'KS'	'DC'	58,176018	85,99499	0,66469663

Tabla G.4. Reglas relacionadas con los ítems quebrados o ítems cero

Resultados de la aplicación de las medidas objetivas para encontrar reglas muy fuertes que cumplen la condición $A+C > 1$.

A	->	C	supp	conf	fcert	SuppA	SuppC	SuppA+C
AM	->	AM	69,694	99,468	0,982	0,701	0,700	1,401
TR,KS	->	AR	7,952	99,844	0,998	0,790	0,291	1,081
AM	->	CP	2,137	78,899	0,551	0,694	0,530	1,224
AM	->	DC	30,026	100,000	1,000	0,300	0,932	1,232
AM	->	IC	62,165	99,910	0,997	0,622	0,677	1,299
AM	->	KC	62,165	91,891	0,785	0,677	0,622	1,299
TR,KS	->	KV	7,959	99,922	0,999	0,790	0,378	1,167
KS,AM	->	KV	17,172	99,783	0,997	0,694	0,378	1,072
VM,AM	->	KV	17,371	99,786	0,997	0,697	0,378	1,075
TR,KS	->	KV	18,346	99,764	0,996	0,790	0,378	1,167
KS,VM	->	KV	17,172	99,783	0,997	0,694	0,378	1,072
AM	->	LG	8,536	100,000	1,000	0,481	0,576	1,057
AM	->	PR	2,268	83,716	0,586	0,694	0,607	1,301

AM	->	SM	16,955	99,817	0,782	0,170	0,992	1,161
AM	->	TR	41,663	100,000	1,000	0,417	0,798	1,215
KS,AM	->	UC	14,668	84,412	0,744	0,694	0,390	1,084
AM	->	VM	69,694	99,556	0,985	0,700	0,701	1,401
TR,KS	->	VR	23,354	99,235	0,989	0,790	0,291	1,080
VM,IC	->	KC,AM	47,869	91,087	0,828	0,526	0,481	1,007
AM	->	KS,AM	69,247	98,830	0,962	0,701	0,694	1,395
AM	->	TR,AM	54,920	78,383	0,517	0,701	0,552	1,253
AM	->	AM,DC	64,600	92,197	0,778	0,701	0,649	1,349
AM	->	IC,DC	58,505	94,029	0,836	0,622	0,636	1,258
AM	->	KC,DC	58,505	86,482	0,674	0,677	0,586	1,262
AM	->	SM,DC	29,660	98,779	0,839	0,300	0,924	1,224
AM	->	TR,DC	39,122	93,901	0,761	0,417	0,744	1,161
AM	->	VM,DC	64,600	92,279	0,780	0,700	0,649	1,349
AM	->	AM,IC	48,093	77,294	0,522	0,622	0,525	1,147
AM	->	KS,IC	61,804	99,331	0,980	0,622	0,672	1,294
AM	->	TR,IC	48,615	78,133	0,534	0,622	0,530	1,153
AM	->	VM,IC	48,124	77,344	0,522	0,622	0,526	1,148
AM	->	KC,KS	61,804	91,358	0,773	0,677	0,619	1,295
AM	->	TR,KS	41,663	100,000	1,000	0,417	0,790	1,206
AM	->	AM,SM	69,694	99,468	0,982	0,701	0,700	1,401
AM	->	IC,SM	61,823	99,361	0,980	0,622	0,673	1,295

Tabla G.5. Ejemplos de reglas muy fuertes que cumplen la condición de $A+C > 1$

Ejemplo de reglas muy fuertes cumplen la condición de $A+C < 1$:

A	->	C	supp	conf	fcert	SuppA	SuppC	SuppA+C	-C	->	-A	Supp-C	supp-A
AM,IP	->	KV	3,995	100	1,000	0,174	0,378	0,552	KC	->	AR,IG,UR	0,622	0,021
KC,VR	->	AR	2,901	100	1,000	0,132	0,291	0,423	AM	->	KV,VR	0,700	0,159
AM,IP	->	TR	3,858	100	1,000	0,174	0,798	0,973	TG	->	AR,IG,UR	0,202	0,021
TG,KC	->	AR	2,584	100	1,000	0,136	0,291	0,427	AM	->	TR,KC	0,700	0,487
AM,IP	->	VM	3,995	100	1,000	0,174	0,701	0,875	VR	->	AR,IG,UR	0,291	0,021
VR,DC	->	AR	6,113	100	1,000	0,274	0,291	0,565	AM	->	VM,DC	0,700	0,649
AM,LM	->	VM	17,408	100	1,000	0,174	0,701	0,875	VR	->	AR,LM	0,291	0,062
VR,DC	->	AR	6,113	100	1,000	0,274	0,291	0,565	AM	->	VM,DC	0,700	0,649
AM,UC	->	VM	14,699	100	1,000	0,270	0,701	0,970	VR	->	AR,UC	0,291	0,117
VR,UR	->	AR	6,511	100	1,000	0,065	0,291	0,356	AM	->	VM,UR	0,700	0,126
AM,UM	->	KS,VM	30,492	100	1,000	0,305	0,694	0,999	KV,VM	->	AR,UM	0,219	0,109
KV,IG	->	AR	3,175	100	1,000	0,098	0,291	0,389	AM	->	KC,IG	0,700	0,116
AM,UM	->	TR,VM	30,492	100	1,000	0,305	0,553	0,858	TG,VM	->	AR,UM	0,148	0,109
TG,KC	->	AR	2,584	100	1,000	0,136	0,291	0,427	AM	->	TR,KC	0,700	0,487
AM,UR	->	VM	12,562	100	1,000	0,126	0,701	0,826	VR	->	AR,UR	0,291	0,065
VR,DC	->	AR	6,113	100	1,000	0,274	0,291	0,565	AM	->	VM,DC	0,700	0,649
AR,DC	->	VR	5,784	100	1,000	0,274	0,291	0,565	VM	->	AM,DC	0,701	0,649
VM,LM	->	AM	10,947	100	1,000	0,174	0,700	0,874	AR	->	VR,LM	0,291	0,062
AR,IC	->	VR	3,075	100	1,000	0,142	0,291	0,433	VM	->	AM,IC	0,701	0,525
VM,IP	->	AM	3,119	100	1,000	0,175	0,700	0,875	AR	->	VR,IG,UR	0,291	0,050
AR,IG	->	VR	9,704	100	1,000	0,097	0,291	0,388	VM	->	AM,IG	0,701	0,174
VM,IP	->	AM	3,995	100	1,000	0,175	0,700	0,875	AR	->	VR,IG,UR	0,291	0,050
AR,LM	->	KS,VR	6,163	100	1,000	0,062	0,288	0,350	KV,VR	->	AM,LM	0,159	0,174
KV,VM	->	AM	5,094	100	1,000	0,219	0,700	0,919	AR	->	KC,VM	0,291	0,482
AR,LP	->	TR	6,952	100	1,000	0,070	0,798	0,868	TG	->	AM,LP	0,202	0,115

TG,KC	->	AM	10,841	100	1,000	0,136	0,700	0,836	AR	->	TR,KC	0,291	0,487
AR,SM	->	VR	6,014	100	1,000	0,284	0,291	0,574	VM	->	AM,SM	0,701	0,700
VM,LM	->	AM	10,947	100	1,000	0,174	0,700	0,874	AR	->	VR,LM	0,291	0,062
AR,UC	->	TR	2,858	100	1,000	0,117	0,798	0,915	TG	->	AM,UC	0,202	0,270
TG,VM	->	AM	14,774	100	1,000	0,148	0,700	0,848	AR	->	TR,VM	0,291	0,553
AR,UC	->	VR	4,939	100	1,000	0,117	0,291	0,407	VM	->	AM,UC	0,701	0,270
VM,IP	->	AM	3,995	100	1,000	0,175	0,700	0,875	AR	->	VR,IG,UR	0,291	0,050
AR,UM	->	TR	10,916	100	1,000	0,109	0,798	0,907	TG	->	AM,UM	0,202	0,305
TG,KC	->	AM	10,841	100	1,000	0,136	0,700	0,836	AR	->	TR,KC	0,291	0,487
AR,UM	->	VR	10,916	100	1,000	0,109	0,291	0,400	VM	->	AM,UM	0,701	0,305
VM,IP	->	AM	3,995	100	1,000	0,175	0,700	0,875	AR	->	VR,IG,UR	0,291	0,050
AR,UR	->	VR	6,511	100	1,000	0,065	0,291	0,356	VM	->	AM,UR	0,701	0,126
VM,IP	->	AM	3,995	100	1,000	0,175	0,700	0,875	AR	->	VR,IG,UR	0,291	0,050
KC,UR	->	IC,LG	11,556	100	0,999	0,116	0,403	0,519	IG,LG	->	KV,UR	0,050	0,078
IG	->	KV	9,754	100	0,999	0,098	0,378	0,475	KC	->	IP	0,622	0,226
VR,IG	->	KV	9,754	100	0,999	0,098	0,378	0,475	KC	->	VM,IG	0,622	0,175
KC,UC	->	IC	23,975	100	0,998	0,240	0,677	0,916	IG	->	KV,UC	0,098	0,150
TG,IP	->	KV,SM	3,839	100	0,997	0,038	0,373	0,411	KC,SM	->	TR,IG,UR	0,619	0,023
KC,UM	->	TR,IC	26,634	100	0,997	0,267	0,530	0,797	TG,IC	->	KV,UM	0,146	0,150

Tabla G.6.Reglas muy fuertes que cumplen la condición de $A+C < 1$

Bibliografía

Publicaciones en General

- [AC05] ACNielsen Chile, "ACNielsen Cambios en el Mercado Chileno", 2005
- [AC06] ACTI, "la revolucion digital y el capital humano", 2006.
- [AH00] Hax, Arnoldo C. y Wilde, Dean L. II: "El modelo Delta – Un nuevo marco conceptual para la nueva economía", 2000.
- [AI76] Ansoff, H. Igor (1976): La estrategia de la empresa, Ediciones Universidad de Navarra, Pamplona.
- [AK84] Andrews K. R. (1971), El concept de estrategia de la Empresa, Barcelo, 1984.
- [AL03] Asociación Latinoamericana de Integración, "La brecha digital y sus repercusiones en los países miembros de la ALADI", 2003.
- [AN00] Norma Alvarez L., María Luisa Aspe A., et al, "el reto de la globalización para la industria mexicana",2000.
- [AN03] Agrawal Narendra and Stephen A. Smith (2003), "Optimal Retail Assortments for Substitutable Items Purchased in Sets," Naval Research Logistics, Vol. 50, pp. 793-822.
- [AT93] R. Agrawal, T. Imielinski, and A. Swami. Mining associations between sets of items in massive databases. In ACM-SIGMOD International Conference on Data, pages 207–216, 1993
- [AP04] Memoria Anual 2004 PARIS
- [AP06] APACS, The definitive guide for the media on plastic card fraud and measures to prevent it, 2006 (<http://www9.secure-ssl-server.com/cardwatch/images/uploads/publications/Fraud%20the%20Facts%202006.pdf>)2
- [AP98] Almaguer Prado Pedro D., Identificación de los parámetros y variables más importantes en el manejo y optimización de una empresa de distribución, PradoLand Consultores, 1998.
- [AR93] Agrawal R., Mielinski T., and A.Swami:" Mining association rules between set of items in large databases", In Proc. Of the 1993 ACM SIGMOD Conference, pages 207-216, 1993.
- [BA05] Baños Ariel, pricing: la nueva herramienta para la rentabilidad empresarial, www.fijaciondeprecios.com
- [BB05] Bouchon Bernardette: "Fuzzy logic Approaches to data mining and information retrieval", Seminario Nuevas Tendencias en Sistemas Inteligentes y Soft Computing, Universidad de Granada, 2005.
- [BC05] Bci Corredor de Bolsa S.A., "análisis de empresa Cencosud S.A. / Sector: Retail", 6 de Diciembre de 2005.
- [BC06] Bci Corredor de Bolsa S.A., "análisis de empresa S.A.C.I. Falabella / Sector: Retail", 3 de Abril de 2006.
- [BF01] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila.,A new framework to assess association rules. In Hoffmann, F., Ed., Adv. in Intelligent Data Analysis. 4th Intern. Symp., IDA'01. Lecture Notes in Comp.Sc. 2189, pag.95–104. Springer-Verlag, 2001.
- [BF02] F. Berzal, I. Blanco, D. Sánchez, and M.A. Vila., Measuring the Accuracy and Interest of Association rules: A New Framework. Intelligent Data Analysis 6 pp. 221–235, 2002.

- [BF03a] F. Berzal, I. Blanco, D. Sánchez, J.M. Serrano, and M.A. Vila. A definition for fuzzy approximate dependencies. Aceptado para su publicación en Fuzzy Sets and Systems.
- [BF03b] F. Berzal, J.C. Cubero, D. Sánchez, J.M. Serrano, M.A. Vila. Finding Fuzzy Approximate Dependencies within STULONG Data. Proceedings of the ECML/PKDD 2003, Workshop on Discovery Challenge. Pp 34–46. 22–26 September 2003.
- [BI00] I. Blanco, M.J. Martín-Bautista, D. Sánchez, M.A. Vila. On the support of dependencies in relational databases: Strong approximate dependencies. Data Mining and Knowledge Discovery. Sometido.
- [BL00] Brause,R; Langsdorf,T; Hepp,M.:“Neural Data Mining for Credit Card Detection”.
- [BM02] Bennett,M.; “How to foil the online fraudsters”, 2002 .
- [BP01] P. Bosc, D. Dubois, O. Pivert, H. Prade. On fuzzy association rules based on fuzzy cardinalities., FUZZY-IEEE, 2001.
- [BP03] Bhatla Tej Paul, Prabhu Vikram, Dua Amit:“Understanding Credit Card Frauds”, 2003.
- [BP83] P.D. Bra & J. Paredaens. Horizontal decompositions for handling exceptions to functional dependencies. Advances in Database Theory, 2:123–144, 1983.
- [BR00] Bolton, Richard J.; Hand, David J. :“Unsupervised Profiling Methods for fraud detection”.
- [BL00] Brause,R; Langsdorf,T; Hepp,M.:“Neural Data Mining for Credit Card Detection”.
- [BR03] Brath, R.; Brody, A:“ Finding the needle in the haystack: using data visualization to spot patterns and anomalies in business data”, 2003.
- [BR99] R. Bayardo and R. Agrawal. Mining the most interesting rules. In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, pages 145–154, 1999.
- [BT03] Bhatla Tej Paul, Prabhu Vikram, Dua Amit:“Understanding Credit Card Frauds”, 2003.
(http://www.tcs.com/0_whitepapers/htdocs/credit_card_fraud_white_paper.pdf).
- [BT04] Baudel,T.; Haible,B.; Sander,G. : “Visual data mining with ILOG discovery”, 2004.
- [CC01] Cámara Chilena de Santiago, “La economía digital en Chile: Primera Parte la Revolución Digital”, 2001.
- [CS04] “5th Annual online fraud report Credit Card Fraud Trends and Merchants' Response” 2004 Edition, Sponsored by CyberSource Corporation Conducted by Mindwave Research.
- [CC04] Chuang-Cheng-Chiu; Chieh-Yuan-Tsai: “A Web services-based collaborative scheme for credit card fraud detection”, 2004.
- [CC06] CCS “Empresas del sector retail proyectan inversiones por más de US\$3000 Millones,2006.
- [CC06] CCS: “Retail en Chile: Respuesta al nuevo consumidor chileno”, 2006.
- [CE03] Castillo, Enrique; Gutierrez, José Manuel; Hadi, Ali S.:“Sistemas Expertos y Modelos de Redes Probabilísticos”, 2003.
- [CG02] G. Chen, Q. Wei. Fuzzy association rules and the extended mining algorithms. Information Sciences 147(1-4): 201–228, 2002.

- [CI97] Cruz Roche, Ignacio y Oubiña Barbolla, Javier (1997): "La Distribución en Europa: Diferentes Modelos", en Díez de Castro, Enrique Carlos (Coordinador) (1997): Distribución Comercial, 2ª edición, McGraw-Hill, Madrid, pp. 139-150.
- [CJ04] J. Calero, G. Delgado, M. Sánchez-Marañón, D., Sanchez, J.M. Serrano, M.A. Vila. An Experience in Mngment of Imprecise Soil DB by Means of Fuzzy Ass. Rules and Fuzzy Approx. Dependencies. ICEIS 2004, 14-17 Abril 2004, Oporto (Portugal).
- [CJ87] Casares Ripol, Javier (comp.); Briz, Julián; Rebollo Arévalo, Alfonso y Muñoz Gallego, Pablo (1987): La Economía de la Distribución Comercial, Editorial Ariel, Barcelona.
- [CJ91] Casares Ripol, Javier y Rebollo Arévalo, Alfonso (1991): "Distribución Comercial. La Aceleración del Cambio. 1966-1991. Veinticinco Años de Revolución Comercial en España", Distribución y Consumo, nº 1, diciembre, pp. 10-38
- [CJ95] J.C. Cubero, J.M. Medina, O. Pons, M.A. Vila. The generalized selection: An alternative way for the quotient operations in fuzzy relational databases. Fuzzy Logic and Soft Computing, B. Bouchon-Meunier, R. Yager, and L.A. Zadeh, Eds. World Scientific Press, 1995.
- [CJ96a] Casares Ripol, Javier y Rebollo Arévalo, Alfonso (1996a): Distribución Comercial, Editorial Cívitas, Madrid.
- [ChK97] K.C.C. Chan, W.-H. Au. Mining Fuzzy asociation Rules. CIKM 1997: 209-215, 1997.
- [CM00] Cahill, Michel H; Lambert, Diane; Pincheiro, Jose; Sun, Don: "Detecting fraud in the real world".
- [CM99] Comité de Marketing-Merchandising de AECOC , » Recomendación sobre surtido eficiente », Octubre 1999.
- [CP00a] Chan, Philip K.: "Toward Scalable Leaning with Non-Uniform class an d cost Distribution: A Case Study in Credit Card Fraud Detection".
- [CP00b] Chan, Philip K.; Fan, Wei; Prodromidis, Andreas; Stolfo, Salvatore J.: "Distributed Data Mining in Credit Card Fraud Detection".
- [CP06] Cuesta Valiño, P. . "Estrategias de crecimiento de las empresas de distribución comercial", Tesis doctoral, 2006.
- [CS03] Cybersource, 5th Annual on line fraud report, 2003.
- [CS04] Memoria Anual 2004 Cencosud
- [CS06] CyberSource Corporation < 7th Annual On Line Fraud Report, On line Payment Fraud Trends Merchant Practices & Benchmarks, 2006 Edition.
- [DD03] D. Dubois, E. Hüllermeier, H. Prade. A Note, on Quality Measures for Fuzzy Association Rules., Fuzzy Sets and Systems - IFSA 2003, LNAI 2715, 346-353, 2003.
- [DD04] Durán J. Dunia: "la dirección estratégica del capital humano con base en el conocimiento, análisis de conocimientos en empresas textiles y de la confección en españa", U.Barcelona, 2004.
- [DF03] Díez, F.J.: "Introducción al Razonamiento Aproximado", 2003.
- [DM00] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "Fuzzy cardinality based evaluation of quantified sentences", International Journal of Approximate Reasoning. Vol 23, pp 23-66, 2000.
- [DM02] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "Acquisition of Fuzzy Association Rules from Medical Data", 2002.
- [DM02a] Delgado Miguel: "Notas sobre aprendizaje automático -Curso 2001/2002".

- [DM03] Delgado Miguel, Sánchez Daniel, Vila María Amparo., Marín Nicolás : "Fuzzy Association Rules: General Model and Application", IEEE Transactions on Fuzzy System 11 (2), pp 214-225, 2003.
- [DM03a] Delgado Miguel, Sánchez Daniel, Vila María Amparo: Mining Fuzzy Association Rules: An Overview. 2003 BISC International Workshop on Soft Computing for Internet and Bioinformatics", accepted. 2003.
- [DM03b] Delgado Miguel, Sánchez Daniel, Vila María Amparo: Fuzzy Association Rules: General Model and Applications. IEEE Transactions on Fuzzy Systems,11(2), pp. 214-225, 2003.
- [DM97] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "Un enfoque lógico para calcular el grado de cumplimiento de sentencias con cuantificadores lingüísticos, 1997.
- [DM98] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "Un método para la evaluación de sentencias con cuantificadores lingüísticos, 1998.
- [DM99a] Delgado Miguel, Sánchez Daniel, Vila María Amparo:"Fuzzy Quantified Dependencies in Relational Databases", 1999.
- [DM99b] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "A logic based definition of fuzzy cardinality, 1999.
- [DM99c] Delgado Miguel, Sánchez Daniel, Vila María Amparo: "Fuzzy cardinality based evaluation of quantified sentences, 1999.
- [DO97] Duarte, Oscar G. "UNFUZZY - Software para el análisis, diseño, simulación e implementación de Sistemas de Lógica Difusa". M.Sc. Tesis. Universidad Nacional de Colombia, Facultad de Ingeniería, Maestría en Automatización Industrial. 1997.
- [DS88] Davis, Stanley M.: Futuro Perfecto, Folio, Barcelona, 1988.
- [DW03] The Data Warehousing Institute, resultado encuesta 2003 a 510 compañías
- [EP02] Essex,P.; Schauer,P.: "Automated fraud detection", 2002.
- [FA97] Fernández Nogales, Ángel (1997): Investigación de Mercados: Obtención de Información, Cívitas, Madrid.
- [FB00] Frenkel, Brian:"Using Artificial Intelligence to Detect Fraud In Credit Cards",
- [FB04] Memoria Anual 2004 Falabella
- [FJ02] Faughnan, John G.:" International Net-Baset Credit Card/Check Card Fraud with Small Charges", 2002.
- [FR03] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: Empresas Almacenes Paris S.A., Paris ", Septiembre 2003.
- [FR04] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: S.A.C.I. Falabella (Falabella)", Septiembre 2004.
- [FR05] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: Tiendas por Departamento en Chile ", Junio 2005.,
- [FR05a] Feller Rate, Standard & Poor's Credit Markets Services: "informe de clasificacion riplely chile s.a.", Septiembre 2005.
- [FR06] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: S.A.C.I. Falabella", Septiembre 2006, <http://www.fitchratings.cl/Upload/A-falabe0906.pdf>
- [FR06a] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: Empresas La Polar S.A. (La Polar)", Junio 2006.

- [FR06b] Fitch Ratings Corporate America Latina, "Sector Comercio/Chile, Análisis de Riesgo: S.A.C.I. Falabella", Mayo 2006.
- [FU96] Fayyad U.M., Piatetshy-Shapiro G, Smyth P: "From Data Mining to Knowledge Discovery: An Overview", Advances in Knowledge Discovery and Data Mining, AAA/MIT Press 1996.
- [GA01] A. Gyenesei. Interestingness Measures for Fuzzy Association Rules. PKDD 2001: 152–164, 2001.
- [GA04] Grupo Acción Digital: "2004-2006 Agenda Digital, te acerca al futuro", 2004.
- [GJ01] J.M. de Graaf, W.A. Kusters, J.J.W. Witteman., Interesting Fuzzy Association Rules in Quantitative, Databases. PKDD 2001: 140–151, 2001.
- [GJ99a] Galindo, José: "Tratamiento de imprecisión en Bases de Datos Relacionales" (pag 40), 1999.
- [GJ99b] Galindo, José: "Tratamiento de imprecisión en Bases de Datos Relacionales" (página 44), 1999.
- [GJ99c] Galindo, José: "Tratamiento de imprecisión en Bases de Datos Relacionales" (página 48), 1999.
- [GJ99d] Galindo, José: "Tratamiento de imprecisión en Bases de Datos Relacionales" (página 134), 1999.
- [GM00] Goebel, Michael; Gruenwald, Le: "A Survey of datamining and knowledge software tools".
- [GP78] Green, Paul E. y Tull, Donald S. (1978): Investigaciones de Mercado, editorial Prentice/Hall Internacional, Londres.
- [HA01] Hax Arnoldo; Wilde Dean: "The Delta Project", Palgrave, 2001.
- [HC02] Hyun Ch.; Shaoning P.; Hong M.; Daijin K.; Sung Y.: "Pattern classification using support vector machine ensemble", 2002.
- [HC04] Hyun-Ch.; Shaoning P.; Hong M.; Daijin K.; Sung Y.: "Constructing support vector machine ensemble", 2004.
- [HE01] E. Hullermeier. Implication-Based Fuzzy Association Rules. PKDD 2001: 241–252, 2001.
- [HJ00] Hernández O. José, Ramírez Q. Ma. José, Ferri R. César: "Introducción a la minería de datos, Pearson Prentice Hall, 2000.
- [HS02] Hawkins, S.; He, H.; Williams, G.; Baxter, R.: "Outlier detection using replicator neural networks", 2002.
- [HS04] Hua S.; Hong Z.; Gui R.: "Applying data mining to detect fraud behavior in customs declaration", 2004.
- [HT99] T.P. Hong, C.S. Kuo, & S.C. Chi. Mining association rules from quantitative data, Intelligent Data, Analysis 3 :363–376, 1999.
- [ID02] IDC Chile y Departamento de Estudios Cámara Nacional de Comercio: "Estudio de Inversión en Tecnologías de la Información (TI), 2002.
- [IT06] IT Governance Global Status Report 2006.
- [JF99] Jiménez Felipe: Lecciones sobre las falacias lógicas. Notas del IES (Institute for the international Education of Students), U.S.A., 1999.
- [JG97] Jonson G, Aholes K: Dirección Estratégica. Análisis de la estrategia de las organizaciones. Prentice – Hall, tercera edición, Madrid, 1997.
- [JK03] Jungwon K.; Ong A.; Overill, R, "Design of an artificial immune system as a novel anomaly detector for combating financial fraud in the retail sector", 2003.

- [JK04] Juca,K.; Boukerche,A.; Sobral,J.; Notare,M.: "Human immune anomaly and misuse based detection for computer system operations: part II", 2004.
- [JL00] López Jesús Alfonso, introduccion y conceptos basicos,
- [JR05] Jarufe, Rina M.; "Tiendas por Departamentos en Chile", Fitchratings, Pag. 1-7, 2005.
- [JT02] Jian T.; Zhixiang Ch.; Ada W.; Cheung, D.: "Enhancing effectiveness of outlier detections for low density patterns", 2002.
- [KC98] C.-M. Kuok, A. Fu, & M. H. Wong. Mining fuzzy association rules in databases, SIGMOD Record 27(1):41-46, 1998.
- [KF02] F. Kerlinger. Investigación del Comportamiento, Mc GrawHill, 3ra. Edición, 2002.
- [KJ95] J. Kivinen, H. Mannila. Approximate Dependency Inference from Relations. Theoretical Computer Science 149(1), pp. 129-149, 1995.
- [KK96] Kalyanam, Kirthi, James S. Hodges and Daniel S. Putler (1996), "A Bayesian Approach to Estimating Target Market Potential with Limited Geo-Demographic Information," Journal of Marketing Research, (1996), Vol XXXIII, (May), pp. 34-49.
- [KM98] M. Kaya, R. Alhaji, F. Polat, A. Arslan. Efficient, Automated Mining of Fuzzy Association Rules. DEXA 2002: 133-142, 2002.27(1):41-46, 1998.
- [KP00] Kotler, Philip; Cámara, Dionisio; Grande, Ildelfonso y Cruz Roche, Ignacio (2000): Dirección de Marketing, Edición del Milenio (10ª edición), Prentice Hall, Madrid.
- [KP99] Kotler, Philip; Armstrong, Gary; Saunders, John; Wong, Verónica; Miquel, Salvador; Bigné, Enrique y Cámara, Dionisio (1999): Introducción al Marketing, 2ª edición europea, Prentice Hall, Madrid.
- [KS98] Klebe Steven W.: "Evaluating on line credit fraud fraud witch artificial intelligence", 1998. (<http://www.wilsonweb.com/wct1/980615ai-screen.htm>)
- [LA04] Lanza, Richard B. ,"How to Use a New Computer Audit Fraud Prevention and Detection Tool", 2004.
- [LB97] Bing Liu, Wynne Hsu, and Shu Chen. Using general impressions to analyze discovered classification rules. In KDD, pages 31-36, 1997.
- [LB98] Bing Liu, Wynne Hsu and Yiming Ma. Integrating Clasdification and Association Rule Mining. In KDD-98, New York, Aug 27-31, 1998.
- [LB99] Bing Liu, Wynne Hsu, Lai-Fun Mun, and Hing-Yan Lee. Finding interesting patterns using user expectations. IEEE Transactions on Knowledge and Data Engeneering, Vol 11(6):817-832, 1999
- [LB00] Bing Liu, Wynne Hsu, Shu Chen, and Yiming Ma. Analyzing the subjective interestingness of association rules. IEEE Intelligent Systems, 15(5):47-55, 2000
- [LE02] Lundin,E.; Kvarnstrom,H.; Jonsson,E.: "A synthetic fraud data generation methodology", 2002.
- [LJ00] J. Luo, S. Bridges. Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection. International Journal of Intelligent Systems, 15(8): 687-703, 2000.
- [LJ87] J.H. Lee & H.L. Kwang. An extension of association rules using fuzzy sets. In Proc. of IFSA'97, Prague, Czech Republic, 1997.

- [LJ94] Lamo de Espinosa, Jaime y Fiel Asenjo, Carmen (1994): "Situación y Perspectivas del Sector de Distribución Alimentaria", Papeles de Economía Española, nº 60/61, pp. 302-315.
- [LJ95] Lambin, Jean-Jacques (1995): Marketing Estratégico, 3ª edición, McGraw-Hill, Madrid.
- [LM04] Lucarini, M.A: "Fighting fraud with technology [insurance fraud detection]", 2004.
- [LP02] Lucas Peter: "Fighting Fraud at the front-end", 2002 (<http://www.collectionsworld.com/08sr02.htm>).
- [LP02] Lucas Peter: "Fighting Fraud at the front-end", 2002 (<http://www.collectionsworld.com/08sr02.htm>).
- [LP03] Memoria Anual 2003 LaPolar
- [LP04] Memoria Anual 2004 La Polar
- [LR03] Landry, R; Debreceny,R; Gray,G: "Grab your picks and shovels! There's gold in your data",2004.
- [MA04] Maxion,A.; Townsend,T.: " Masquerade detection augmented with error analysis",2004.
- [MC99] Comité de Marketing-Merchandising de AECOC, "RECOMENDACIÓN SOBRE SURTIDO EFICIENTE", 1999.
- [ME75] E.H. Mamdani, J.J. Odtengaard, E. Lembessis. "An experiment in Linguistic synthesis with a fuzzy logic controller". Int J. Man. Machine Studies.; Vol 7 págs:1-13; 1975
- [ME83] E.H. Mamdani, J.J. Odtengaard, E. Lembessis. "Use of fuzzy logic for implementing rule-based control of industrial processes". Advances in Fuzzy Sets.; Possibility Theory and Applications (Wang eds) Plenum Press 1983.
- [MK97] Malhotra, Narres K. (1997): Investigación de Mercados, Un Enfoque Práctico, 2ª edición, Prentice Hall, México.
- [ML02] MacVittie,L.: "Online fraud detection takes diligence",2002.
- [ML04] Min J.; Taek S.: " A neural classifier with fraud density map for effective credit card fraud detection", 2004.
- [ML93] Masa Godoy, Leopoldo (1993): "Perfil de las Formas Comerciales. Elementos de la Tecnología de Venta", Distribución y Consumo, nº10, junio-julio, pp. 20-33.
- [MN06] Management & Negocios en la Nueva Economía," PRICING: UN JUEGO DE PERCEPCION".
- [MN95] Muñiz Martínez, Norberto (1995): Alternativas de Diversificación e Internacionalización y Evolución de los Grupos Estratégicos en el Comercio Minorista Europeo y Español. Tesis doctoral, Editorial Universidad de León - Servicio de Publicaciones, León.
- [MN98] Muñiz Martínez, Norberto; Placer Galán, José Luis y Cervantes Blanco, Miguel (1998): "Diversificación e Internacionalización de las Empresas de Distribución Comercial que Operan en España", Distribución y Consumo, nº 37, diciembre 97/enero 98, pp. 58-67.
- [MR00] Morillas Raya Antonio, Introducción al análisis de datos difusos.
- [MS96] Miquel Peris, Salvador; Parra Guerrero, Francisca; L´hermie, Christian y Miquel Romero, Mª José (1996): Distribución Comercial, ESIC Editorial, Madrid.
- [MS97] Miquel Peris, Salvador; Bigné, Enrique; Lévy, Jean-Pierre; Cuenca, Antonio Carlos y Miquel, Mª José (1997): Investigación de Mercados, McGraw-Hill, Madrid.
- [MS99] McIntyre, Shelby H. and Christopher M. Miller (1999) "The Selection and Pricing of Retail Assortments: An Empirical Approach," Journal of Retailing, Vol. 75, No. 3, pp. 295-318.
- [NN00] Detección de fraude en tarjetas de crédito con redes neuronales

- [OA04] Oliviera,A.; Neto,F.; Meira,S: "Combining MLP and RBF neural networks for novelty detection in short time series", 2004.
- [PA07] Parisi Antonio, Rodriguez Arturo, Rodriguez Cristian: "Tecnicas Avanzadas para la predicción de la variación del Ishare MSCI Brazil Indez (EWZ)", Estudios de información y control de gestión, U.Chile, 2007.
- [PA99] Prodromidis, Andreas L.; Stolfo, Salvatore: "Agent-Based Distributed Learning Applied to Fraud Detectión", 1999.
- [PB95] B. Pfahringer & S. Kramer. Compression-based,evaluation of partial determinations. In Proc. First Int'l Conf. Knowledge Discovery and Data Mining (KDD'95), pages 234-239, 1995.
- [PG02] Pérez Hoyos Gustavo, Sistemas de Lógica Difusa, 2002
- [PG91] G. Piatetsky-Shapiro and W. Frawley. Discovery, Analysis and presentation of strong rules in: Knowledge Discovery in Dataabase, AAAI/MIT Press, pp 229-238, 1991.
- [PG94] G. Piatetsky-Shapiro and C. J. Matheus. The interestingness of deviations. In Proceedings of AAAI Wordshop on Knowledge Discovery in Databases, 1994.
- [PM01] Porter, Michael, Artículo Harvard Business Review, Marzo, 2001
- [PM79] Porter, Michael E. (1979): "The Structure Within Industries and Companies Performance", Review of Economics and Statistics, vol. 14, p. 214-227.
- [PM80] Porter, Michael E. (1980): Competitive Strategy: Tecchniques for Analyzing Industries and Competitors, The Free Press, New York; traducido al castellano por Vasseur Walls, Alfonso (1987): Estrategia Competitiva: Técnicas para el Análisis de los Sectores Industriales y de la Competencia, CECSA, Méx.
- [PM92] Porter, Michael. (1986): Estrategia Competitiva. Técnicas para el Análisis de los Sectores Industriales y de la Competencia, C.E.C.S.A., 8ª impresión, México, 1992.
- [PM92a] Porter, Michael. (1987): Ventaja Competitiva. Creación y Sostenimiento de un desempeño superior, C.E.C.S.A., 8ª impresión, México, 1992.
- [PR98] Perry, Raúl:"Detección de Fraude utilizando una Base de Datos Relacional y lógica Difusa", 1998.
- [PW98] W. Pedrycz. Fuzzy Set Technology in Knowledge Discovery. Fuzzy Sets and Systems 98, pp.279-290, 1998.
- [RC04] Rong-Chang Chen, Ming-Li Chiu, Ya-Li Huang, Lin-Ti Chen: "Detecting Credit Card Fraud by using Questionnaire-Responded Transaccion Model Bases on Support Vector Machines", 2004.
- [RD00] Ribeiro, D. : "Gestión competitiva de los recursos humanos en redes de innovación". Harvard Deusto Business Review, nº 95, pp. 77-79.
- [RD03] Roussinov D., and Zhao J., Automatic discovery of similarity relationships through web minin. Decision support system, 35:149-166,2003.
- [RI05a] Román Moreno Ignacio Rafael: "revenue management en retail: estimación de demanda", Departamento de Ingeniería Industrial y Sistemas, Escuela de Ingeniería Pontificia Universidad Católica de Chile, 2005
- [RI05b] Román Moreno Ignacio Rafael: "Beneficios financieros de mejorar el GMROI", Departamento de Ingeniería Industrial y Sistemas, Escuela de Ingeniería Pontificia Universidad Católica de Chile, 2005
- [RJ01] Roddick J. and Rice S, What's interesting about cricket? -on thresholds and anticipation in discovered rules. SIGKDD Exploration,

- 3(1):1-5, 2001.
- [RJ03] Riedl,J.; Hill,R.: "Proceedings of the Fifteenth Innovative Applications of Artificial Intelligence Conference",2003.
- [RM07] Ruiz J. Maria D: "Hacia una Formalización Lógica del Concepto de Interés para Reglas de Asociación", 2007.
- [RP05] Memoria Anual 2005 Ripley
- [RR00] Restrepo Rivas, Luis Guillermo Introducción al aprendizaje de la Maquina: Sistemas Discente".
- [RT90] Robinson, T.M. y Clarke-Hill, C.M. (1990): "Directional Growth by European Retailers", International Journal of Retail & Distribution Management, vol. 18, nº 5, pp. 3-15.
- [SD00] Sánchez. Daniel; Sánchez J.R; Serrano J.M.; Vila, Amparo: "Reglas de Asociación sobre Dominios Imprecisos Aplicando Relaciones Difusas de Similitud".
- [SD99] Sánchez Fernandez Daniel: "Adquisición de relaciones entre atributos en base de datos relacionales", PhD thesis, Departamento Of computer Science and Artificial Intelligenc, Univerity of Granada, December 1999.
- [SE75] E. Shortliffe & B. Buchanan. A model of inexact reasoning in medicine. Mathematical Biosciences, 23:351–379, 1975.
- [SJ03] Serrano, José María, et al: "Knowledge fusion in relational Database:Agregation and Summarization Measures", 2003.
- [SM00] Sam, Maes; Tuys, Kart; Vanschoenwinkel,Bram; Manderick, Bernad: "Credit Card Detection Using Bayesian and Neural Networks".
- [SM02] Syeda,M.; Yan Q.; Yi P.: " Parallel granular neural networks for fast credit card fraud detection",2002.
- [SM03] Szczepanski,M.: "Unsupervised fraud selection", 2003.
- [SM96] Santesmases Mestre, Miguel (1996): Términos de Marketing. Diccionario-Base de Datos, Editorial Pirámide, Madrid.
- [SM99] Santesmases Mestre, Miguel (1999): Marketing. Conceptos y Estrategias, 4ª edición, Ediciones Pirámide, Madrid.
- [SR95] R. Srikant, R. Agrawal. Mining Generalized Association Rules. Procs. of the 21st Int'l Conference on Very Large Databases. Zurich, Suiza, Septiembre, 1995
- [SR96] Srikant R. and Agrawal. R. Mining quantitative association rules in large relational tables. In Proc. 1996 ACM SIGMOD Int'l. Conf. Management Data, Pages 1-12, 1996.
- [SR96] R. Srikant, R. Agrawal. Mining Quantitative Association Rules in Large Relational Tables. Procs.of the ACM SIGMOD Conference on Management of Data, 1996.
- [SS00] Stolfo, Salvatore; Fan, David W.; Lee, Wenke; Prodromidis, y Andreas L. "Credit Card Detection Using Meta-Learning Issues and Initial Results".
- [SS00a] Stephen A. Smith , Achabal Dale D., Shelby H. McIntyre and Kirthi Kalyanam (2000) "A Decision Support System for Vendor Managed Inventory," Journal of Retailing, Vol. 76, No. 4, pp. 430-454.
- [SS01] S. Sahar. Interestigness preprocessing. In Proceedings of the 2001, IEEE Internacional Conference on Data Mining, pages 489–496, San Jose, CA, 2001.
- [SS02] S. Sahar. On incorporating subjective interestingness into the mining process. In Proceedings of the 2002, IEEE International Conference

- on Data Mining, pages 681–684, Maebashi City, Japan, 2002.
- [SS96] Stephen A. Smith and Agrawal Narendra (1996), "Estimation Methods for Retail Inventory Management with Unobservable Lost Sales," *Naval Research Logistics*, Vol. 43, pp.839-861
- [TA90] Toffler, Alvin: *El cambio del poder*, Plaza & Janés, Barcelona, 1990.
- [TA91] Tordjaman, A. y Dionisio, A. (1991): "Estrategias de Internacionalización de las Empresas Comerciales", Colección de Estudios Comercio y Distribución, Comisión de las Comunidades Europeas, Dirección General XXIII, Junio, Bruselas.
- [TE00] Telefónica España: "La Sociedad de la Información en España. Situación actual y perspectivas", 2000.
- [TE94] Trillas, Enric, et al, "Fundamentos e Introducción a la Ingeniería FUZZY, Omrom", 1994.
- [UA02] Angélica Urrutia S; Marcela Varas C., José Galindo G, "Diseño de una base de datos difusa Modelada con UML, 2002.
- [UC06] INAP, Universidad de Chile, Brecha Digital y Sociedad de la Información: *Paradojas y Nuevos Paradigmas*, 2006
- [VA02] Vila, Amparo: "Modelos de datos avanzados: Bases de datos orientadas a objetos", 2002.
- [VJM05] Vivanco VonBorries José Miguel: "Revenue management: Diagnostico y evaluacion de oportunidades en cadena de retail", Departamento de Ingeniería Industrial y Sistemas, Escuela de Ingeniería Pontificia Universidad Católica de Chile, 2005
- [VR97] Vázquez Casielles, Rodolfo y Trespalacios Gutiérrez, Juan Antonio (1997): *Distribución Comercial. Estrategias de Fabricantes y Detallistas*, Editorial Cívitas, Madrid.
- [VS02] Viaene,S.; Derrig,R.; Dedene,G.: "Boosting naive Bayes for claim fraud diagnosis",2002.
- [VS04] Viaene,S.; Derrig,R.; Dedene,G: "A case study of applying boosting naive Bayes to claim fraud diagnosis", 2004.
- [WA97] Wilensky, Alberto: *Política de negocios, Estrategias de Marketing para mercados competitivos*, Ediciones Macchi, Buenos Aires, 1997.
- [WD04] Wang-Dong; Wang-Quan-yu; Zhan-Shou-yi; Li-Feng-xia; Wang-Da-zhen; " A feature extraction method for fraud detection in mobile communication networks", 2004.
- [WH98] W.H. Au & K.C.C. Chan. An effective algorithm,for discovering fuzzy rules in relational databases. In *Proc. of IEEE International Conference on Fuzzy Systems*, vol II, 1314–1319, 1998
- [WM02] Weatherford,M.: " Mining for fraud", 2002.
- [WS99] S.-L. Wang,J.-S. Tsai, B.-C. Chien. Mining Approximate Dependencies Using Partitions on Similarity-relation-based Fuzzy Databases. *Proceedings of the IEEE SMC'99*, October 1999, Tokyo, Japan, V-871–V-875.
- [XL04] Xideris,L.M.: "Claiming the future [claims processing]", 2004.
- [YF95] Y. Fu and J. Han. Meta-rule-guide of asociation rules in relational database. In *Prec. First Int. Workshop Integration of KKD with deductive and Object-Oriented databases (KD00D'95)*.
- [YH02] H.-T. Yang, S.-J. Lee. Mining Fuzzy Association Rules from Sequence databases with Quantitative data and Inter-Transaction Intervals. *FSKD 2002*: 606–610.
- [YK04] Yamanishi,-K.; Takeuchi,J.; Williams,G.; Milne,P. "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms", 2004.

- [YK04] Yufeng-Kou; Chang-Tien-Lu; Sirwongwattana,S.; Yo-Ping-Huang: "Survey of fraud detection techniques", 2004.
- [YT88a] M.M.Gupta, T.Yamakawa (eds.), "Fuzzy Logic in Knowledge-Based Systems, Decision and Control", North-Holland / Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
- [YT88b] M.M.Gupta, T.Yamakawa (eds.), "Fuzzy Computing: Theory, Hardware, and Applications", North-Holland / Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1988.
- [ZL65] L. A. Zadeh. "Fuzzy Sets". Information and Control; Vol 8 págs:338-353; 1965
- [ZL71] L. A. Zadeh. Similarity Relations and Fuzzy Orderings. Information Sciences, 3:177-200, 1971.
- [ZL73] L. A. Zadeh. "Outline of a new approach to the analisis of complex systems and decision processes". IEEE: Trans. Syst., Man, Cybern.; Vol 3 págs:28-44; January-1973
- [ZL83] L. A. Zadeh. A computational approach to fuzzy quantifiers in natural languages. Computing and Mathematics with Applications, vol. 9, no. 1, pp. 149-184, 1983.
- [ZL88] L. A. Zadeh. "Fuzzy Logic". IEEE Computer Mag; págs:83-93; Apr-1988
- [ZZ97] Z. Zhang, Y. Lu, B. Zhang. An Effective Partitioning-Combining Algorithm for Discovering Quantitative Association Rules. National Natural Science Foundation of China, 1997.

Publicaciones Periodísticas

- [DF03] Diario Financiero, artículo: Los cambios que provocará la fusión, 9 de junio de 2003.
- [EL04a] Diario El Mercurio, Economía y Negocios: artículo: Cuantas Tarjetas hay en su Billetera, 19 de Marzo de 2004.
- [EL04b] Diario El Mercurio, artículo: "Y no ardió Paris", Santiago de Chile, 22 de Agosto de 2004.
- [EL05] Diario El Mercurio, Economía y Negocios, artículo: " DISPUTA. Holding de los Luksic busca frenar la fusión de Paris con Cencosud, La encrucijada de Quiñenco, por María Eugenia Tamblay C., 6 de marzo de 2005.
- [EM04] Diario El Mercurio, Economía y Negocios: artículo: 91 grandes grupos empresariales existen en Chile, según la SVS, 20 de febrero de 2004.
- [EM04a] Diario El Mercurio, Economía y Negocios: artículo: Cuantas Tarjetas hay en su Billetera, 19 de Marzo de 2004.
- [EM06] Diario El Mercurio, Artículo 28 de Mayo de 2006, El desafío de Lazaro Calderon, 2006.
- [LT04] Diario La Tercera, artículo: " El clan que ahora remeció al comercio", Santiago de Chile, 22 de Agosto de 2004.
- [LT05] Diario La Tercera, artículo: " Una agitada semana da paso a días decisivos en la confrontación: Luksic y Paulmann miden fuerzas y planes de negocio en Almacenes Paris", Fecha edición: 06-03-2005, Carola Rojas y Víctor Cofré.

Sitios Web de Interés

- [01] <http://ciberconta.unizar.es/LECCION/REDES/172.htm>
- [02] <http://www.aicorp.co.uk/risknet.asp>
- [03] http://www.cybersource.com/products_and_services/credit_card_fraud_management/fraud_screening/

- [04] <http://www.electronica.com.mx/neural/informacion/backpropagation.html>
- [05] <http://www.eng.salleurl.edu/~se04184/P2Definitions.html>
- [06] <http://www.fraudintelligence.com/index2.html>
- [07] http://www.infovis.net/Revista/num_31.htm [
- [08] http://www.lightbridge.com/global/sp/behavior_monitor.html
- [09] <http://www.lsi.us.es/~joaquinp/doc/fuzzy.html>
- [10] http://www.neuralt.com/nt2/products/minotaur_transcure1.htm
- [11] <http://www.uniovi.es/~constitucional/fundamentos/primero/bastidan.htm>
- [12] http://www.infovis.net/Revista/num_31.htm
- [13] <http://www.uniovi.es/~constitucional/fundamentos/primero/bastidan.htm>
- [14] <http://www.lsi.us.es/~joaquinp/doc/fuzzy.html>
- [15] <http://www.eng.salleurl.edu/~se04184/P2Definitions.html>
- [16] http://www.cybersource.com/products_and_services/credit_card_fraud_management/fraud_screening/
- [17] [http://www.fitchratings.cl/BajarArchivo.asp?Archivo=retailjun05\(1\).pdf2](http://www.fitchratings.cl/BajarArchivo.asp?Archivo=retailjun05(1).pdf2)
- [18] <http://www.novedadesenred.com/v2/pages/numero.asp?ArtID=1846#pr>
- [19] http://www.tcs.com/0_whitepapers/htdocs/credit_card_fraud_white_paper.pdf
- [20] <http://www.eumed.net/tesis/2006/pcv/>
- [21] http://members.tripod.com/jesus_alfonso_lopez/FuzzyIntro.html
- [22] <http://luisguillermo.com/discentes.pdf>.