

CLASIFICACIÓN DE CLIENTES MEDIANTE TÉCNICAS DE MINERÍA DE
DATOS PARA LA EMPRESA COMERTEX S.A.

CARLOS ALBERTO RIVERO SARMIENTO

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE ESTUDIOS INDUSTRIALES Y EMPRESARIALES
BUCARAMANGA

2012

CLASIFICACIÓN DE CLIENTES MEDIANTE TÉCNICAS DE MINERÍA DE
DATOS PARA LA EMPRESA COMERTEX S.A.

CARLOS ALBERTO RIVERO SARMIENTO

Trabajo de Grado para optar por el título de:
INGENIERO INDUSTRIAL

Director:

HENRY LAMOS DIAZ

Ph. D en Matemáticas

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
ESCUELA DE ESTUDIOS INDUSTRIALES Y EMPRESARIALES
BUCARAMANGA

2012

DEDICATORIA

*A Luz, Martha y Juan padres como ningunos,
a Javier, Liliana y Ricardo por compartir este sueño.*

AGRADECIMIENTO

A Henry por orientar y compartir su conocimiento. Por catar afectos con paciencia.

Al grupo de investigación *ÓPALO* por generar conocimiento, por la disposición y disponibilidad de los recursos para la ejecución y cumplimiento de cada uno de los objetivos.

A COMERTEX S.A. por facilitar e impulsar el desarrollo de este trabajo, a cada uno de los trabajadores que aportaron su experiencia y tiempo en el desarrollo del mismo.

TABLA DE CONTENIDO

INTRODUCCIÓN	18
1. GENERALIDADES DE LA EMPRESA.....	20
1.1 DESCRIPCIÓN DE LA EMPRESA	20
1.1.1 Objeto social.	20
1.1.2 Reseña histórica.	20
1.1.3 Estructura organizativa.	21
1.2 DIRECCIONAMIENTO ESTRATÉGICO	21
1.2.1 Misión.	21
1.2.2 Visión.	21
1.2.3 Razón de ser.	22
1.3 PRODUCTOS	22
1.4 SEGMENTACIÓN DE CLIENTES.....	23
1.4.1 Materia Prima.....	23
1.4.2 Producto Terminado.	25
2. GENERALIDADES DEL PROYECTO	27
2.1 PLANTEAMIENTO DEL PROBLEMA.....	27
2.2 JUSTIFICACIÓN DEL PROBLEMA	28
2.3 OBJETIVOS.....	29
2.3.1 Objetivo general.....	29
2.3.2 Objetivos específicos	29
2.4 ALCANCE	30
3. MARCO TEÓRICO	31
3.1 DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS (<i>KDD</i>)	31
3.2 MINERÍA DE DATOS.....	35
3.2.1 Técnicas de la MD.	37
3.2.2 Minería de datos en los negocios.	56
3.2.3 Minería de datos y su relación con el CRM.	57

3.3 METODOLOGÍAS PARA PROCESOS DE MINERÍA DE DATOS	61
3.4 PROBLEMAS COMBINATORIOS EN LA MINERÍA DE DATOS	62
4. PROCESO DE DESCUBRIMIENTO EN BASES DE DATOS	64
4.1 INTEGRACIÓN Y RECOPIACIÓN.....	64
4.2 SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN.....	65
4.2.1 Depuración de la base de datos.	65
4.3 MINERÍA DE DATOS.....	79
4.3.1 Redes Bayesianas en SPSS Modeler.....	79
4.3.2 Árboles de decisión en SPSS.	85
4.3.3 Análisis de clúster.	88
5. RESULTADOS	89
5.1 MINERÍA DE DATOS Y LEVANTAMIENTO DE INFORMACIÓN	89
5.2 PROMESA DE SERVICIO	99
5.2.1 Asesor comercial.	100
5.2.2 Producto.	102
5.2.3 Entregas.	103
5.2.4 Atención quejas y reclamos.	103
5.2.5 Información.	104
5.3 Estrategias de servicio por segmento	104
5.4 CLIENTES VIP	113
5.5 BASE DE DATOS	115
CONCLUSIONES.....	117
RECOMENDACIONES	119
BIBLIOGRAFÍA	121
ANEXOS	125

LISTA DE TABLAS

Tabla 1. Distribución de CEDI's	21
Tabla 2. Líneas de producto por UEN	22
Tabla 3. Datos de los clientes para el 2011	42
Tabla 4. Datos de entrenamiento para la clasificación de clientes VIP	51
Tabla 5. Industrias y áreas de aplicación de la MD	57
Tabla 6. Prefijos plantilla 10	67
Tabla 7. Bodegas que registra SEVEN-ERP	68
Tabla 8. Bodegas consideradas para el proyecto	69
Tabla 9. Segmentos plantilla 10	70
Tabla 10. Campos eliminados plantilla 10	70
Tabla 11. Códigos de Bodega	73
Tabla 12: Valores del nivel de satisfacción	90
Tabla 13. Probabilidades condicionales del Asesor Comercial	91
Tabla 14: Participación de los segmentos	97
Tabla 15: Criterios evaluados al Asesor Comercial	100
Tabla 16. Modelo para el registro de pedidos	102
Tabla 17: Comparación de segmentos	105
Tabla 18: Comparación de segmentos	106
Tabla 19: Tipos de clientes	113
Tabla 20: VIP por segmentos	114

LISTA DE FIGURAS

Figura 1. Relación DATO-INFORMACIÓN-CONOCIMIENTO	32
Figura 2. Proceso KDD	32
Figura 3. Tareas de la MD.....	36
Figura 4. MD y su relación con otras áreas del conocimiento	36
Figura 5. Red Bayesiana	39
Figura 6. Red construida	43
Figura 7. Esquema de un Árbol de decisiones	46
Figura 8. Árbol de decisiones (parcial)	52
Figura 9. Fases de la metodología CRISP-DM	62
Figura 10. Paleta de nodos de origen SPSS Modeler	79
Figura 11. Nodo origen de datos sobre el lienzo	80
Figura 12. Datos de entrada nodo origen	80
Figura 13. Nodo tipo.....	81
Figura 14. Conexión nodos origen-tipo	81
Figura 15. Personalización nodo tipo	81
Figura 16. Selección del papel de cada variable	82
Figura 17. Nodo modelado de red bayesiana.....	82
Figura 18. Parámetros para el nodo red bayesiana	83
Figura 19. Ejecución de la ruta.....	83
Figura 20. Nugget resultado de la ejecución de la ruta	84
Figura 21. Resultados del modelo de red bayesiana	84
Figura 22. Ruta para cargar base de datos a SPSS PASW	85
Figura 23. Ruta para crear el árbol de decisiones.....	86
Figura 24. Definir variables del modelo	86
Figura 25. Ventana para definir criterios del árbol.....	87
Figura 26. Criterios de definición del árbol	87
Figura 27. Red bayesiana para satisfacción del cliente	90
Figura 28. Red bayesiana de líneas de producto	91
Figura 29. Importancia del predictor.....	92

Figura 30. CART	93
Figura 31. Tamaño de los conglomerados	95
Figura 32. Importancia de los predictores para la construcción de conglomerados.....	95
Figura 33. Infografía caracterización de un grupo de clientes	98
Figura 34: Relación causas de insatisfacción.....	99
Figura 35: Evaluación promedio a los criterios asociados al Asesor Comercial	100
Figura 36: Comparación del aspecto evaluado y la importancia que tiene para el negocio	101

LISTA DE ANEXOS

Anexo A: ESTRUCTURA ORGANIZACIONAL COMERTEX S.A.	125
Anexo B. PLANTILLAS DE TRABAJO Y ENCUESTA DEL NIVEL DE SERVICIO	126
Anexo C. MANUAL DE CODIFICACIÓN.....	127
Anexo D. DEFINICIÓN DE LÍNEAS DE PRODUCTO.....	133
Anexo E. CLASIFICACIÓN DE PRODUCTOS POR GRUPO.....	135
Anexo F. MODELOS DE REDES BAYESIANAS	141
Anexo G. CLASIFICACIÓN DE CLIENTES MEDIANTE CART	144
Anexo H. CARACTERIZACIÓN DE SEGMENTOS.....	147
Anexo I. PROMESA DE SERVICIO	149

TABLA DE CUMPLIMIENTO DE OBJETIVOS

OBJETIVO	DESCRIPCIÓN	CUMPLIMIENTO
1	Descubrir mediante clasificadores de Redes Bayesianas relaciones de independencia o relevancia entre las variables observadas.	Capítulo 4 y 5
2	Identificar a través de los árboles de decisión la clasificación de clientes adecuada para COMERTEX S.A.	Capítulo 4 y 5
3	Obtener grupos o conjuntos entre los datos mediante el análisis de clúster de acuerdo a su asociación/similitud.	Capítulo 4 y 5
4	Identificar oportunidades/necesidades no atendidas en cada uno de los segmentos.	Capítulo 5
5	Construir el modelo de clasificación de clientes para COMERTEX S.A.	Capítulo 5

GLOSARIO

ALGORITMO: Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema.

ANÁLISIS DE CLÚSTER: Tarea de minería de datos que trata de encontrar grupos entre un conjunto de individuos u objetos a través de las características que estos poseen.

ÁRBOLES DE DECISIÓN: Es una estructura que visualmente describe una serie de condiciones que conducen a la toma de una decisión.

CART: Es un algoritmo de árbol de decisión que realiza particiones binarias sobre los atributos, pero que va asignando una media y una varianza a cada nodo, intentando seleccionar las particiones que reduzcan las varianzas de los nodos hijos.

CRM: *Customer Relationship Management*, estrategia de negocios centrada en el cliente.

CUBOS OLAP: *Procesamiento analítico en línea (OLAP, on-line analytical processing)*, el cual facilita un análisis multidimensional y ofrece una presentación a diferentes niveles de abstracción.

ERP: *Enterprise Resource Planning*, software que integra todos los procesos de un organización.

HEURÍSTICA: Técnica que aumenta la eficiencia de un proceso de búsqueda, posiblemente sacrificando algunos requisitos de optimalidad, es decir, no garantiza la mejor respuesta pero casi siempre una buena solución.

IBM SPSS MODELER: Es un área de trabajo de minería de datos que permite elaborar modelos predictivos de forma rápida e intuitiva, sin necesidad de programación. Descubre patrones y tendencias en datos estructurados o no estructurados mediante una interfaz visual soportada por análisis avanzado. Modela resultados y muestra los factores que influyen en ellos. Así, aprovecha las oportunidades y atenúa riesgos.

KDD: *Knowledge Discovery in Databases*, descubrimiento de conocimiento en bases de datos. Se refiere al proceso no trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información.

METAHEURÍSTICA: Método aproximado diseñado para resolver problemas de optimización combinatoria en los que los heurísticos clásicos no son efectivos. Proporcionan un marco general para nuevos algoritmos híbridos, combinando diferentes conceptos de la inteligencia artificial, mecanismos estadísticos y otros.

MINERÍA DE DATOS: Descubrimiento automático de patrones o modelos interesantes y no obvios ocultos en una base de datos, los cuales tienen un gran potencial para contribuir en los aspectos principales del negocio.

PROBLEMAS COMBINATORIOS: Es una rama de la investigación operativa que consiste en encontrar la solución óptima a un problema en que cada solución está asociada a un determinado valor (valor de la solución).

REDES BAYESIANAS: Es un grafo dirigido acíclico que consiste en un conjunto de variable aleatorias que construyen los nodos de la red, un conjunto de arcos dirigidos que conectan pares de nodos y para cada nodo se especifica su función de distribución condicional que cuantifica los efectos que los padres tienen sobre el nodo.

RESUMEN

TÍTULO: CLASIFICACIÓN DE CLIENTES MEDIANTE TÉCNICAS DE MINERÍA DE DATOS PARA LA EMPRESA COMERTEX S.A.[†].

AUTOR: CARLOS ALBERTO RIVERO SARMIENTO[‡]

PALABRAS CLAVES: Minería de datos, Redes bayesianas, Árboles de decisión, Análisis de clúster, Clientes, CRM.

DESCRIPCIÓN

La minería de datos surge como resultado de la necesidad de trabajar con grandes volúmenes de datos sin que se pierda la escalabilidad, principal limitación de la estadística básica. La importancia de la minería de datos para las empresas radica en que es un medio de búsqueda de información no visible de los clientes, y de esta manera orientar las estrategias hacia ellos de una forma más eficiente. En la presente investigación se muestra la bondad de la minería de datos para el caso de una empresa ubicada en el departamento de Santander y que pertenece al sector textil. Se usan herramientas como cubos OLAP, árboles de regresión, redes bayesianas y análisis de clúster con el propósito de encontrar patrones de los clientes y de esta manera construir perfiles que ayuden a la empresa en la toma de decisiones. Se usa el software de IBM SPSS Modeler como herramienta computacional para la construcción de las técnicas. Luego del análisis y evaluación de los datos obtenidos, se logró caracterizar los clientes elaborando una tipología que permite mejorar el servicio a ellos. Se construyeron protocolos para captura de datos y la información se presenta mediante infografías que resumen criterios relevantes para cada área de la empresa y facilitan la toma de decisiones soportadas en datos actualizados.

Este trabajo es presentado como instrumento de difusión del proyecto de investigación modalidad **Cofinanciación, convocatoria Colciencias No. 502 de 2010. COD. 1102-502-27270 - CT. 727-2011.**

[†] Proyecto de grado

[‡] Facultad de Ingenierías Físico-Mecánicas, Escuela de Estudios Industriales y Empresariales, Director Ph D Henry Lamos Diaz

ABSTRACT

TITLE: CLASSIFICATION OF CUSTOMERS USING DATA MINING TECHNIQUES FOR COMPANY COMERTEX S.A.[§]

AUTHOR: CARLOS ALBERTO RIVERO SARMIENTO**

KEY WORDS: Data mining, Bayesian networks, Decision trees, Cluster analysis, Customer, CRM.

DESCRIPTION

The data mining starts as a result of the need about working with large volumes of data without losing scalability, the main limitation of basic statistics. The data mining importance for companies lies in that it is a way to search information about clients that is not visible thus, it can be oriented strategies to them in a more efficient way. In this investigation, it is showed the benefits of data mining with a company that is in Santander, Colombia and belongs to the textile field. Tools like OLAP cubes, regression trees, Bayesian networks and cluster analysis in order to find patterns about clients and create profiles that help the company to make decisions. The IBM SPSS modeler software is used as a computational tool for building those techniques. After the analysis and evaluation of the data obtained the clients were characterized making a typology that allows improving the service to them. Were made protocols for data catching and the information is presented through info graphics that summarize relevant facts for each area in the company and make easier everything about making decisions that are based on actualized data.

This research is an instrument of diffusion of the investigation project: confinanciación, convocatoria colciencias No.502 de 2010.cod. 1102-502-27270- CT. 727-2011.

[§] Work degree

^{**} Faculty of Physico-Mechanica Engineering, School of Industrial and Business Studies, Director Ph D Henry Lamos Diaz

INTRODUCCIÓN

El incremento en la cantidad y el tipo de datos que actualmente almacenan las empresas ha propiciado el desarrollo de nuevas técnicas de estudio para el beneficio de las mismas, buscando extraer nueva información de los diferentes procesos que facilite la toma de decisiones.

Todo ha apuntado a la necesidad de metodologías de análisis inteligente de datos que logre la generación de conocimiento útil a partir de los datos históricos y así predecir el desempeño futuro. Es por esto que la búsqueda de nuevos productos y servicios, acordes con los nuevos mercados, y soportados por un sistema de información transaccional permite a COMERTEX S:A. realizar un estudio, mediante la MINERÍA DE DATOS, para clasificar los clientes, encontrar hábitos de consumo y clientes potenciales a partir de diferentes patrones que puedan ser explícitos en los datos y sobre los cuales se pueden crear estrategias de comercialización o servicio con valor agregado que permitan la diferenciación del servicio.

La información en cualquier organización reduce la incertidumbre sobre algún aspecto de la realidad y, por tanto, permite tomar buenas decisiones y aprovechar las oportunidades de “negocio”. La clasificación de los clientes ha proporcionado un enfoque diferente a las relaciones entre cliente y empresa, a través de características o similitudes, facilitando la construcción de estrategias de comercialización y relaciones más próximas que generen beneficios en las dos direcciones. La aplicación de herramientas de minería de datos en la gestión de las relaciones con el cliente es una tendencia en la economía mundial. Analizar y comprender los comportamientos de los clientes y sus características es la base del desarrollo de una estrategia competitiva con el fin de adquirir y retener a los clientes potenciales y maximizar el valor del cliente.

El objetivo del presente trabajo está enmarcado desde la construcción de un modelo conceptual que permita la elaboración de perfiles de los clientes mediante la identificación de diferentes relaciones de independencia/dependencia entre las variables de estudio así como la identificación de oportunidades de mejora producto del análisis comparativo de los datos.

La estructura del inicia con el conocimiento acerca de la empresa, las generalidades del proyecto, el marco teórico en relación al conocimiento de la minería de datos y su utilidad en beneficio de la gestión de las relaciones con el cliente. Posterior a esto, el trabajo de grado muestra la ejecución metodológica que siguen los proyectos que utilizan la minería de datos y finaliza con la especificación de resultados y estrategias planteadas para la construcción de los modelos conceptuales así como sus respectivas conclusiones, recomendaciones y bibliografía.

1. GENERALIDADES DE LA EMPRESA

1.1 DESCRIPCIÓN DE LA EMPRESA

1.1.1 Objeto social.

COMERTEX S.A. cuenta con 6 líneas de productos con el objeto social de adquirir, exportar, importar, fabricar, transformar, distribuir y vender toda clase de insumos y productos textiles. Cuenta con dos unidades estratégicas de negocio (UEN), materia prima y producto terminado, que atiende trece segmentos del mercado textil. Adicional a esto distribuye algodón, principal insumo para las textileras del país.

1.1.2 Reseña histórica.

El 7 de noviembre de 1.974 fue fundada en la ciudad de Bucaramanga; el surgimiento de esta empresa se vio motivado por la necesidad que percibieron sus creadores de fortalecer uno de los sectores más ricos de la economía del país, la distribución de productos textiles, visualizando en Santander una de las mejores plazas para aprovechar dicha oportunidad.¹

COMERTEX S.A. es una empresa consolidada como un grupo de empresas familiares que se fusionaron hacia el año 2004, en las ciudades de Bogotá, Cali y Pereira. En la actualidad cuenta con una moderna infraestructura ubicada en la zona industrial de Girón en la Carrera 17 Autopista Palenque – Chimitá N. 60– 170 en la que funcionan sus líneas comerciales, departamento administrativo y su centro de distribución logístico. En esta sede los productos de las líneas son almacenados en una bodega amplia con las características logísticas necesarias, con el fin de brindar a los clientes mayor confiabilidad y un manejo seguro de los productos. Además posee, en los principales puntos

¹ Suministrado por COMERTEX S.A.

comerciales de la ciudad de Bucaramanga, almacenes donde funcionan sus siete líneas, logrando así satisfacer la gran demanda generada en el sector.²

1.1.3 Estructura organizativa.

COMERTEX S.A. cuenta con 250 personas vinculadas directamente y ubicadas en 7 gerencias. (Ver Anexo A: Estructura organizacional).

En la actualidad, COMERTEX S.A. cuenta con 3 centros de distribución ubicados en diferentes ciudades del país. (Ver Tabla 1)

Tabla 1. Distribución de CEDI's

CEDI	CIUDAD	CAPACIDAD (m³)
Oriente	Girón	8991
Centro	Bogotá	2513
Occidente	Pereira	1440

Fuente. COMERTEX S. A.

1.2 DIRECCIONAMIENTO ESTRATÉGICO

1.2.1 Misión.

“Somos una sociedad anónima, cerrada, que realiza negocios rentables, que promueve el desarrollo de nuestros empleados y colaboradores, que fideliza a nuestros clientes y proveedores, que responde a los objetivos de los accionistas y que contribuye al mejoramiento de la calidad de vida de la sociedad.”³

1.2.2 Visión.

“Al finalizar el año 2.012 seremos comercializadores orientados al cliente, como una sociedad anónima, abierta a alianzas y socios; estaremos representando, comercializando productos terminados con marcas propias y de terceros, produciendo en terceros materias primas y productos intermedios con valor

²IBID

³IBID

agregado, tanto a nivel nacional e internacional, brindando a proveedores y clientes una infraestructura moderna.”⁴

1.2.3 Razón de ser.

“COMERTEX S.A., una familia donde accionistas, clientes, empleados y proveedores se apoyan mutuamente para crecer de manera sostenida, atendiendo segmentos y canales de la cadena textil, con insumos y productos competitivos para calzado, hogar y prendas de vestir.

El desarrollo del talento humano, la actualización de los recursos tecnológicos y el mejoramiento continuo de los procesos impulsa nuestra dinámica de cambio e innovación constante, donde la premisa básica es crecer fundamentados en comercialización de talla mundial con servicio cálido, ágil y confiable.”⁵

1.3 PRODUCTOS

El portafolio de productos que comercializa está distribuido en dos unidades estratégicas de negocio (UEN) y dividido en 6 líneas de producto. (Ver Tabla 2)

Tabla 2. Líneas de producto por UEN

UNIDAD ESTRATÉGICA DE NEGOCIO	
Materia Prima	Producto Terminado
No Tejido	Vestuario
Tejido Punto	Hogar
Tejido Plano	Calzado

Fuente. COMERTEX S.A.

La identificación y características de las líneas de producto comercializadas se relacionan a continuación:

⁴IBID

⁵IBID

- Tejido Plano: En esta línea se encuentran diferentes tipos de tela para la confección de prendas de vestir, tanto casuales como deportivas; de igual manera ofrece telas para lencería, ropa de trabajo e industrial.
- Tejido Punto: La línea cuenta con la siguiente variedad de telas: deportivos, desagujados, doble punto, estampados, lycras, perchados, piqué, preteñidos, punto sencillo, rib y camisetas.
- No Tejidos: Ofrece productos para diferentes sectores de la industria como son acolchadores, colchoneros, bondeadores, confección, calzado y marroquinería, filtración (telas no tejidas), impermeabilización, agrícola, publicidad y empaque, higiénico y quirúrgico.
- Hogar: Los productos que se manejan son: toallas y sus subgéneros; ropa de hogar y productos importados.
- Calzado: Los productos de esta línea se clasifican en cinco grandes grupos de acuerdo a su uso en: doméstico, agropecuario, industrial, colegial y moda.
- Vestuario: Esta línea cuenta con productos tales como: camisetas, pantalones, vestidos, camisas, pantalonetas, entre otros. Maneja también marcas propias y franquicias de reconocidas tiendas de ropa, tales como Naf Naf y Chevignon.

1.4 SEGMENTACIÓN DE CLIENTES

En la actualidad, COMERTEX S.A. ha identificado 13 segmentos de mercado con características particulares, por ejemplo, de acuerdo a la Unidad Estratégica de Negocio y al objeto social. (Ver Tabla 3).

1.4.1 Materia Prima.

- Comercio Mayorista: Pertenecen a este segmento los clientes que ejercen las actividades de venta, representación y promoción de los productos de materias primas para ser revendidos. Distribuyen y venden principalmente al por mayor y en ocasiones al detal. Estos clientes pueden tener sus propios puntos de venta. Principales características: poco o nada fieles, interés en productos de línea, alto volumen en

pedidos, favorecen la disponibilidad, la exclusividad en diseños, la atención personalizada, precio y plazos de pago.

Tabla 3. Segmentos por UEN

UNIDAD ESTRATÉGICA DE NEGOCIO	
Materia Prima	Producto Terminado
Comercio Mayorista	Grandes Superficies
Comercio Minorista	
Confecciones Dotaciones	
Confecciones Femenina	Mayoristas
Confecciones Infantil	
Confecciones Masculina	
Confecciones Quirúrgica	
Confecciones Ropa deportiva	Comercio al detal
Industrial	
Publicidad	

Fuente. COMERTEX S.A.

- Comercio Minorista: Son establecimientos comerciales pequeños donde sus clientes pueden adquirir productos de materias primas al detal. Generalmente están especializados en pocos productos y atienden clientes con necesidades particulares. Principales características: clientes de bajo volumen en pedidos, orientados a productos de línea, con gran exigencia en tiempos de entrega. Orientados a buscar descuentos.
- Confecciones: Clientes dedicados a la elaboración de prendas de vestir. Este segmento se encuentra dividido en: dotaciones, infantil, femenina, masculina, quirúrgica y ropa deportiva. Principales características: clientes especializados, favorecen una atención personalizada con buen

nivel de asesoría técnica, privilegian diseño, y son exigentes en programación y entrega.

- Industrial: Empresas cuya actividad primordial es la producción de bienes mediante la transformación de la materia o extracción de materias primas. Principales características: clientes especializados, favorecen una atención personalizada con buen nivel de asesoría, exigentes en diseño, programación y entrega.
- Publicidad: Clientes dedicados principalmente a la comercialización de productos terminados y/o materias primas para fines publicitarios. Prestan éste servicio a empresas interesadas en diseños de productos personalizados que exhiban sus logos o marcas. Principales características: clientes de pedidos de bajo volumen; favorecen la atención personalizada y la disponibilidad de productos con buen precio, exigen cumplimiento en fechas de entrega.

1.4.2 Producto Terminado.

- Grandes Superficies: Son tiendas únicas de gran tamaño, ubicadas en las principales ciudades del territorio nacional, dentro de las cuales están hipermercados, almacenes de cadena, supermercados, almacenes por departamentos. Principales características: logística especial (etiquetado, reempaque, despachos a varios puntos de envío), demanda de publicidad (impulso, mercaderistas, góndolas).
- Mayoristas: Son clientes que compran para distribuir e invierten en la comercialización de los productos terminados; estos son: catálogo de venta directa, institucional, promocional y licitaciones. Principales características: requieren logística básica con valor agregado escaso o nulo, este segmento maneja su propia logística para llegar al cliente final. Comercializan y distribuyen el producto.

- Comercio al detal: Es el segmento que finaliza la cadena de distribución de los productos terminados, venden directamente al cliente final y su sistema de venta es al detal. Principales características: demandan alta rentabilidad porque sus despachos son pequeños y deben cubrir los costos logísticos de operación. Proceso logístico con nivel alto de complejidad.

2. GENERALIDADES DEL PROYECTO

2.1 PLANTEAMIENTO DEL PROBLEMA

COMERTEX S.A. identificó que sus clientes se diferencian altamente en la compra según el objeto social que desempeñan; en este sentido, busca entender a los clientes y su comportamiento para ser más asertivos en la toma de decisiones relacionadas con la gestión de los clientes y en la forma en cómo se deben atender sus necesidades.

La orientación hacia el producto y un conocimiento reducido de los clientes ha limitado la toma de decisiones asertivas que garanticen un aumento en la participación del mercado. En este sentido, COMERTEX S.A. encuentra necesario la generación de un nuevo enfoque dirigido a los clientes a través del conocimiento profundo de sus necesidades.

La búsqueda de nuevos productos y servicios, acordes con los nuevos mercados, y soportados por un sistema de información transaccional permite realizar un estudio, mediante la MINERÍA DE DATOS (Data Mining), para clasificar los clientes, encontrar hábitos de consumo y clientes potenciales a partir de diferentes patrones que puedan ser explícitos en los datos y sobre los cuales se pueden crear estrategias de comercialización con valor agregado que permita la diferenciación del servicio.

La información en cualquier organización reduce la incertidumbre sobre algún aspecto de la realidad y, por tanto, permite tomar buenas decisiones y aprovechar las oportunidades de “negocio”. Esta información pretende generar conocimiento a partir de experiencias reales de adaptación de modelos y réplicas de análisis de clúster, árboles de decisión y redes bayesianas para la re-clasificación de los clientes de COMERTEX S.A. a partir de los segmentos que actualmente considerados.

La clasificación de los clientes ha proporcionado un enfoque diferente a las relaciones entre cliente y empresa, a través de características o similitudes, facilitando la construcción de estrategias de comercialización y relaciones más próximas que generen beneficios en las dos direcciones.

El sistema de información SEVEN-ERP utilizado por COMERTEX S.A. contiene diferentes módulos en las áreas de logística, comercial y administrativa; en la actualidad se cuenta con una base de datos que contiene información que puede resultar valiosa en la generación de nuevos negocios.

La importancia de la ejecución y el éxito de éste proyecto está en la elaboración de perfiles para los diferentes segmentos y generar oportunidades específicas de negocio, además de servir de insumo para el proyecto “DISEÑO DE MECANISMOS Y HERRAMIENTAS CONCEPTUALES Y TECNOLÓGICAS PARA EL MEJORAMIENTO DEL SISTEMA LOGÍSTICO DE LA EMPRESA COMERTEX S.A.”, mediante el uso de técnicas de Minería de Datos para la formación académica e investigativa.

2.2 JUSTIFICACIÓN DEL PROBLEMA

A nivel de COMERTEX S.A., la importancia está en la identificación de oportunidades de mejora evidenciadas en la encuesta del nivel de servicio en cuanto al asesor comercial, producto, entregas y atención telefónica impulsan el cambio de estrategia, es decir, pasar de una estrategia orientada al producto a una estrategia centrada en el cliente que permita nivelar las expectativas que tiene con respecto al servicio junto a lo que percibe de éste.

El aprovechamiento de los avances tecnológicos en el almacenamiento de datos y el desarrollo de diferentes técnicas para extraer información en forma eficiente, para su uso en aplicaciones de negocios, que le permitan reducir la incertidumbre y el costo asociado a la toma de decisiones para la gestión de sus clientes en pro de la creación de estrategias de comercialización o servicio diferenciadas que incentiven la consolidación de la empresa en el mercado.

La importancia de la realización del presente trabajo de grado está dada por la intervención de diferentes actores (Universidad-Empresa-Estado), en el afianzamiento de esta relación, el desarrollo de métodos y herramientas para aumentar el nivel de efectividad en el servicio de COMERTEX S.A. mediante la generación de conocimiento proporcionado por investigación universitaria e impulsada por el estado.

Para la universidad, la adaptación de técnicas, con frecuencia académicas, al ámbito empresarial y el desarrollo de modelos para la generación de nuevos negocios con productos/servicios, que creen valor, destinados a clientes o segmentos específicos a través de la identificación de perfiles, características o tendencias ocultas contenidas en las bases de datos de COMERTEX S.A. mediante la aplicación de minería de datos facilitan la integración y generación de conocimiento.

2.3 OBJETIVOS

2.3.1 Objetivo general

Diseñar un modelo conceptual utilizando técnicas de minería de datos para la construcción de perfiles de los clientes de COMERTEX S.A., que permita a la organización tomar decisiones estratégicas sobre los segmentos actuales de la empresa.

2.3.2 Objetivos específicos

- Descubrir mediante clasificadores de Redes Bayesianas relaciones de independencia o relevancia entre las variables observadas.
- Identificar a través de los árboles de decisión la clasificación de clientes adecuada para COMERTEX S.A.
- Obtener grupos o conjuntos entre los datos mediante el análisis de clúster de acuerdo a su asociación/similitud.
- Identificar oportunidades/necesidades no atendidas en cada uno de los segmentos.

- Construir el modelo de clasificación de clientes para COMERTEX S.A.

2.4 ALCANCE

En la realización de este trabajo de grado se espera obtener la caracterización de los clientes de acuerdo a los segmentos que actualmente se manejan y que permitan la generación de nuevas estrategias de comercialización. El proyecto se llevará a cabo desde una perspectiva **NO DIRIGIDA** en la cual se parte de los datos para la obtención de patrones que, a medida que se van descubriendo, se estime si pueden ayudar a resolver algunas necesidades del negocio.

El proyecto inicia con la consolidación, depuración y adecuación de los datos para el posterior tratamiento estadístico y de minería de datos.

Mediante software como SPSS MODELER y SPSS PASW se aplicaran diversas técnicas de minería de datos como Redes Bayesianas, Árboles de Decisión y Análisis de Clúster que permitan la consolidación de resultados de relaciones entre las variables, clasificación de los clientes y la asociación de clientes en grupos o conjuntos, respectivamente.

Por último, el entregable final es un *modelo conceptual de clasificación de clientes* para la empresa que facilite la generación de nuevos negocios y la toma de decisiones.

3. MARCO TEÓRICO

El nivel elevado de tecnología para el almacenamiento y procesamiento de los datos ha contribuido a que las empresas guarden una gran cantidad de ellos que crecen de manera exponencial. Estas bodegas de datos requieren técnicas mucho más elaboradas que el simple uso de herramientas básicas de estadística, que sólo permiten generar información resumida, poco flexible y, sobre todo, poco escalable a grandes volúmenes de datos. Con el fin de poder extraer conocimiento que se considere valioso para la empresa han aparecido nuevos enfoques más refinados e inteligentes.

3.1 DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS (KDD)

La Minería de Datos (Data Mining) ha surgido por la necesidad de conocer, trabajar y aprovechar el aumento de los datos permitiendo a las compañías concentrarse en la información más importante de sus bases de información.

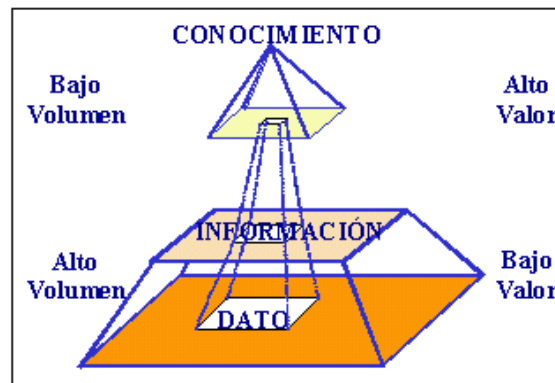
De forma general, los datos son la materia prima bruta. En el momento en que se les atribuye algún significado especial pasan a convertirse en información. Una vez los especialistas elaboran o encuentran un modelo haciendo que la interpretación del confronto entre la información y ese modelo representen un valor agregado entonces nos referimos al conocimiento⁶. En la figura 1 se presenta un modelo de la relación entre datos, información y conocimiento.

En el primer nivel de la pirámide se muestran los datos brutos con los cuales se desea generar nuevo conocimiento para la empresa, conocimiento que se caracteriza por tener un alto valor agregado. La minería de datos trabaja en el nivel superior buscando patrones, comportamientos, agrupaciones, secuencias, tendencias o asociaciones que puedan generar algún modelo que permita

⁶ MOLINA, Luis Carlos. Data Mining: Torturando a los datos hasta que confiesen, 2002.

comprender mejor el dominio de los datos para ayudar en una posible toma de decisiones.

Figura 1. Relación DATO-INFORMACIÓN-CONOCIMIENTO



Fuente. Data Mining: Torturando a los datos hasta que confiesen. Luis Molina. 2002

En su definición, la Minería de Datos se conoce como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos⁷. Es decir, la tarea fundamental de la MD es encontrar modelos inteligibles a partir de los datos y en cuyo proceso se espera descubrir patrones que ayuden a la toma de decisiones más seguras y que generen algún beneficio para la organización. Frecuentemente se utilizan sinónimos para referirse a la MD, por ejemplo, “Descubrimiento de Conocimiento en Bases de Datos” (KDD por sus siglas en inglés), siendo esta última un proceso que consta de una serie de fases y en la cual la MD es una de ellas. (Ver figura 2)

Figura 2. Proceso KDD



Fuente. Introducción a la minería de datos. José Hernández et al, 2004, p.13

⁷ WITTEN & FRANK, Data Mining: Practical Machine Learning Tools and Techniques, citado por HERNÁNDEZ, José et al, Introducción a la minería de datos, 2004, p. 5.

El proceso incluye la selección, limpieza, transformación y proyección de los datos; analizar los datos para extraer patrones y modelos adecuados; evaluar e interpretar los patrones para convertirlos en conocimiento. El conocimiento extraído debe ser válido, novedoso, potencialmente útil y en última instancia comprensible.

La diferencia puntual entre la MD y el KDD radica en que éste último es el proceso global de descubrir conocimiento útil de las bases de datos mientras que la MD se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos, siendo minería de datos el término más utilizado a nivel empresarial por diferentes connotaciones como aventura y dinero fácil⁸.

Fases del Proceso de Extracción de Conocimiento

El proceso *KDD* es iterativo e interactivo. Iterativo ya que la salida de alguna de las fases puede hacer volver a los pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. E interactivo porque el usuario debe ayudar en la preparación de los datos, validación del conocimiento extraído, etc. Las fases del KDD son⁹:

- a. *Fase de integración y recopilación*: En esta fase se determinan las fuentes de información que pueden ser útiles y el origen de ellas, esto es, fuentes de información útiles y dónde conseguirlas. Luego, los datos se transforman a un formato común, frecuentemente mediante un *almacén de datos (Data Warehousing)* que consiga unificar de manera operativa toda la información recogida. Los almacenes de datos se utilizan para poder agregar y cruzar información. Estos almacenes de datos son adecuados para el *procesamiento analítico en línea (OLAP, on-line analytical processing)*, el cual facilita un análisis multidimensional y ofrece una presentación a diferentes niveles de abstracción.

⁸ HERNÁNDEZ, José et al. Introducción a la minería de datos, 2004, p. 22

⁹ IBID

Diferenciar la minería de datos y OLAP es importante; OLAP es un proceso deductivo cuya utilidad radica en comprobar rápidamente patrones y pautas hipotéticas sugeridas por el usuario con el objetivo de verificarlas o rechazarlas mediante la obtención de información agregada a partir de información detallada, en tanto la MD es un proceso inductivo y busca encontrar patrones ocultos.

b. Fase de selección, limpieza y transformación: En esta fase se corrigen o eliminan los datos incorrectos como valores que no se ajustan al comportamiento (*outliers*) o presencia de datos faltantes (*missing values*) y se decide la estrategia a seguir con los datos incompletos. Además, se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes. La calidad del conocimiento obtenido no depende solo del algoritmo que se utilice en la MD, también depende de la calidad de los datos minados, es decir, del buen trabajo realizado al conjunto de datos en esta fase del proceso.

c. Fase de minería de datos: Con esta fase se pretende producir nuevo conocimiento. Se realiza a través de la construcción de un modelo basado en los datos recopilados y el cual es una descripción de los patrones y relaciones encontradas en los datos que pueden usarse para predicciones o para explicar situaciones pasadas. Para esto es necesario tener en cuenta:

- La determinación del tipo de tarea de minería más adecuado.
- La elección del tipo de modelo
- La elección del algoritmo de minería que resuelva la tarea y obtenga el tipo de modelo que se está buscando.

d. Fase de evaluación e interpretación: Aquí se evalúan los patrones y se analizan. Si es necesario se vuelve a las fases anteriores para una nueva iteración. Se espera que los patrones descubiertos tengan tres

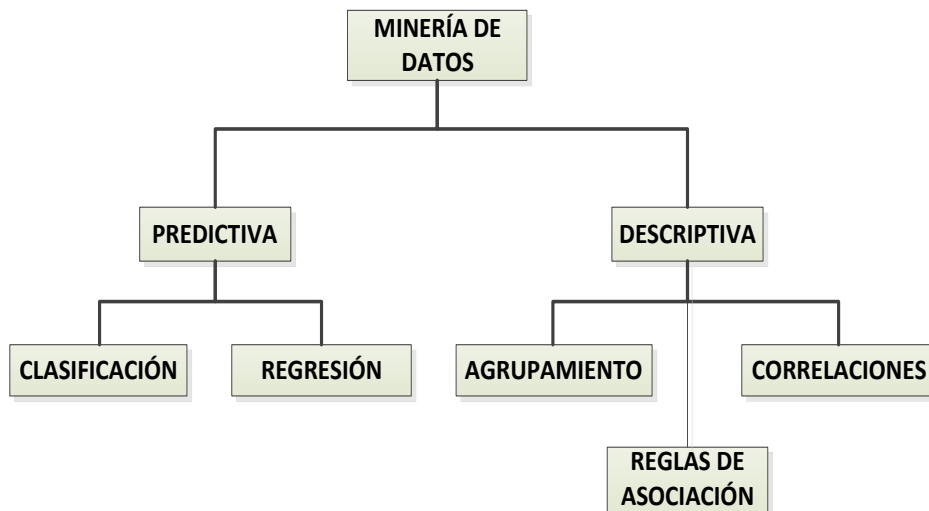
cualidades: precisos, comprensibles (inteligibles) e interesantes (útiles y novedosos). Según las aplicaciones puede interesar mejorar algún criterio y sacrificar ligeramente otro, situación en la cual se regresa a alguna fase anterior del proceso. Pese a las medidas en la evaluación, en muchos casos hay que evaluar el contexto donde se va a utilizar el modelo pues la precisión del modelo no garantiza que refleje el mundo real. En cualquier caso, se debe contrastar el conocimiento que éste proporcione con el conocimiento previo que se puede tener sobre el problema para detectar y en su caso resolver los posibles conflictos.

- e. *Fase de difusión, uso y monitorización*: Se usa el nuevo conocimiento y se hace partícipe de él a todos los posibles usuarios. Es importante medir cómo evoluciona el modelo, pues los patrones pueden cambiar lo cual hace que el modelo deba ser monitoreado, es decir, que cada cierto tiempo el modelo tendrá que ser evaluado, re-entrenado y posiblemente reconstruido completamente.

3.2 MINERÍA DE DATOS

De acuerdo con Peacock [1998], se define la MD como el “*descubrimiento automático de patrones o modelos interesantes y no obvios ocultos en una base de datos, los cuales tienen un gran potencial para contribuir en los aspectos principales del negocio*”. Los modelos de la MD se clasifican en modelos *predictivos* y modelos *descriptivos*. En los primeros se pretende estimar valores futuros o desconocidos de variables de interés usando variables o campos contenidos en la base de datos mientras que en los segundos se identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos. Cada tarea puede ser realizada usando diferentes técnicas, por ejemplo, para la tarea predictiva, clasificación y regresión, se utilizan los árboles de decisión en tanto que para la tarea descriptiva, se utiliza el agrupamiento y reglas de asociación. (Ver Figura 3.)

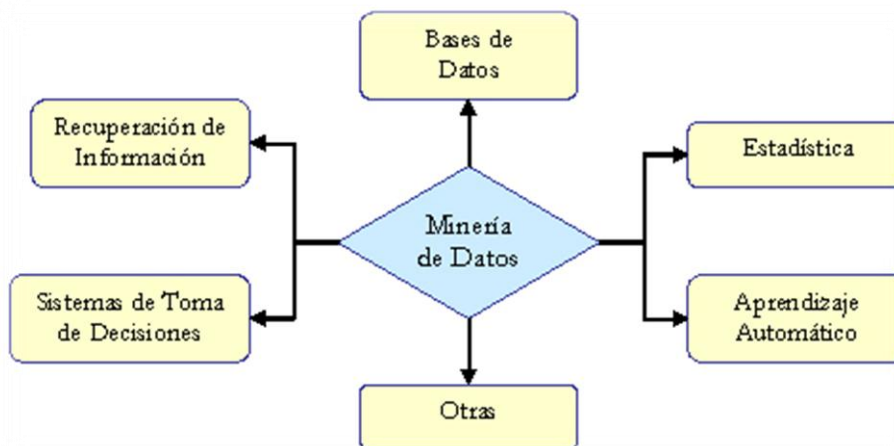
Figura 3. Tareas de la MD



Fuente. Autor

La minería de datos es un área multidisciplinaria debido al uso de diferentes áreas de investigación como el sistema de manejo de bases de datos, inteligencia artificial y el aprendizaje automático. En la figura 4 se muestra la relación de la minería de datos con otras áreas del conocimiento.

Figura 4. MD y su relación con otras áreas del conocimiento



Fuente. Minería de datos, Olmos y González. Instituto Tecnológico de Puebla. 2007.

La importancia y el uso creciente de aplicaciones de MD a nivel empresarial hacen de ella una disciplina de gran auge y constante desarrollo tanto a nivel teórico como práctico. A continuación, se presentan algunas de las ventajas sobre otras herramientas de manejo de datos:

- La MD ayuda a los empresarios en el procesamiento de bodegas de datos para descubrir relaciones que, en algunos casos, ni siquiera se sospechaba.
- La información obtenida a través de la MD ayuda a los usuarios a elegir cursos de acción y a definir estrategias competitivas, porque conocen información que solo ellos pueden emplear.
- Se puede trabajar siguiendo los mismos criterios con grandes cantidades de información histórica.
- El proceso de búsqueda puede ser realizado por herramientas que automáticamente buscan patrones porque así están programadas y despliegan los tópicos más importantes.

3.2.1 Técnicas de la MD.

Como se mencionó anteriormente, existen diversas técnicas y algoritmos ligados a las tareas de la MD, por ejemplo las que aparecen a continuación.

3.2.1.1 Redes Bayesianas (RBs).

Una red bayesiana es una forma de representar el conocimiento cualitativo mediante un grafo dirigido acíclico¹⁰ y un conjunto de probabilidades condicionales que cuantifican las relaciones presentes.

Las RBs ofrecen un análisis cualitativo de los atributos y valores que pueden intervenir en el problema y dan cuenta de la importancia cuantitativa de estos atributos. El aspecto cualitativo hace referencia a la forma en que se representan las relaciones entre los atributos, ya sea en forma casual, o señalando simplemente la correlación que existe entre las variables (o atributos). En la parte cuantitativa, siendo ésta el gran aporte de las RBs, da una medida probabilística de la importancia de cada una de las variables en el problema y, por lo tanto, una probabilidad explícita de las hipótesis que se formulan.

¹⁰ Grafo dirigido acíclico, es un tipo de grafo en el cual sus arcos son dirigidos y no forman ciclos.

El conocimiento se articula en la definición de relaciones de independencia/dependencia entre las variables que componen el modelo. Las relaciones abarcan desde una independencia completa hasta una dependencia funcional entre las variables del modelo. El hecho de utilizar una representación gráfica para la especificación del modelo hace de las RBs una herramienta interesante como representación del conocimiento, aspecto fundamental en la minería de datos.

La teoría de la probabilidad y los métodos bayesianos son técnicas que frecuentemente se usan en la minería de datos. Dos razones caracterizan a los métodos bayesianos:

1. Es un método práctico para realizar inferencias a partir de los datos.
2. Facilita un marco de trabajo útil para la comprensión y análisis de otras técnicas de aprendizaje y minería de datos que no trabajan explícitamente con probabilidades.

Entre las características que poseen las RBs se pueden resaltar¹¹:

- Cada ejemplo observado modifica la probabilidad de que la hipótesis formulada sea correcta (aumenta o disminuye).
- Son métodos robustos al posible ruido presente en los ejemplos de entrenamiento y a la posibilidad de tener entre esos ejemplos de entrenamiento datos incompletos y posiblemente erróneos.
- Permiten tener en cuenta en la predicción de la hipótesis el conocimiento a priori o conocimiento del dominio en forma de probabilidades.

❖ Formulación Matemática

Una red Bayesiana, para un conjunto de variables aleatorias $X = \{X_1, \dots, X_n\}$, es un par $B = (G, P(\Theta))$, donde G es un grafo dirigido acíclico, cuyos nodos se

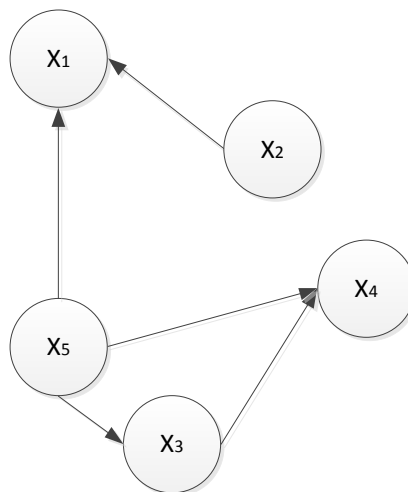
¹¹ MALAGÓN, Constantino. Clasificadores bayesianos. El algoritmo de Naïve Bayes. 2003

relacionan uno a uno con las variables en el vector aleatorio X , y P es un conjunto de funciones de probabilidad local definidas por el conjunto de parámetros Θ . Se usa Pa_i y pa_i para denotar, respectivamente, a los padres y las configuraciones de los padres del nodo X_i en G . La distribución de probabilidad conjunta del vector X puede expresarse como:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i/Pa_i)$$

Un ejemplo de red Bayesiana se presenta en la figura 5.

Figura 5. Red Bayesiana



Fuente. Autor

La función de probabilidad conjunta representada en esta red es:

$$P(X_1, \dots, X_5) = P(X_1/X_2, X_5)P(X_2)P(X_3/X_5)P(X_4/X_3, X_5)P(X_5)$$

Una RB puede ser usada para calcular una probabilidad de interés usando métodos para el proceso de inferencia exacta y aproximada. Este proceso es

de tipo NP-hard¹² (non-polinominal-hard), y se debe a ciclos no dirigidos en la red.

Obtener una red Bayesiana a partir de datos es un proceso de aprendizaje, el cual consta de dos etapas:

- Aprendizaje paramétrico: conocida la estructura de la red, entonces, se debe obtener las probabilidades *a priori* y condicionales requeridas.
- Aprendizaje estructural: consiste en obtener las relaciones de dependencia e independencia entre las variables involucradas de la red Bayesiana.

❖ Estimación de parámetros de una red Bayesiana

En esta sección se supone que la estructura de la red Bayesiana es conocida y que el conjunto de datos está completo. Sea G^h que denota la hipótesis que la función de probabilidad de $X = \{X_1, \dots, X_n\}$ puede factorizarse de acuerdo a la estructura G . Sea $\theta_G = \{\theta_1, \dots, \theta_n\}$ un conjunto donde θ_i es el vector de parámetros de la función de probabilidad conjunta como

$$P(X / \theta_G, G^h) = \prod_{i=1}^n P(x_i / pa_i, \theta_i, G^h)$$

El problema de estimación de parámetros en una red Bayesiana se reduce a calcular la función de probabilidad posterior $P(\theta_G / D, G^h)$, donde D representa el conjunto de datos de entrenamiento. Suponiendo que los parámetros θ_i son mutuamente independientes, entonces:

$$P(\theta_G / D, G^h) = \prod_{i=1}^n P(\theta_i / D, G^h)$$

¹² Non-polinominal-hard, conjunto de problemas de decisión donde los casos no pueden ser decididos en tiempo polinomial por una máquina.

En otras palabras, cada vector de parámetros puede estimarse independientemente de los otros.

❖ Estimación de la estructura de la red Bayesiana

La estructura de la red nos da información sobre las dependencias probabilísticas entre las variables y sus dependencias condicionales dada (s) otra (s) variable (s). Estas dependencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de probabilidades).

Un inconveniente de las redes bayesianas es que requieren que el espacio de las variables aleatorias (los nodos) sean definidos como discretos y se deben conocer las probabilidades a priori. Las RBs realizan la simplificación de que todos los atributos son independientes.

❖ Redes bayesianas como clasificadores

La estimación de redes bayesianas, informalmente, puede describirse como: dada una muestra $D = \{u_1, \dots, u_N\}$ encontrar una red que mejor la describa. La red aprende un modelo que incluye todas las variables (clase y atributos) del problema y luego realiza la clasificación. La metodología es encontrar una función que evalúe cada red con respecto a D y luego seleccione la mejor de acuerdo a esta función. En una red bayesiana se cumple que toda variable es independiente del resto dado su envolvente de markov, definido como:

$$(padres(x) \cup hijos(X) \cup padres(hijos(X)))$$

Es posible seleccionar únicamente dichas variables como variables predictoras útiles para el modelo de clasificación. En general, este problema de optimización no es manejable.

Ejemplo 1: Redes Bayesianas

En una empresa se tiene información de algunos clientes referente a *ubicación*, *segmento*, *tipo de ventas*, *utilidad operacional* y si es un cliente *VIP*. A partir de estos datos se desea construir una red que los describa y las probabilidades a posteriori dadas las dependencias e independencias presentes en la red. Los datos se muestran en la tabla 3:

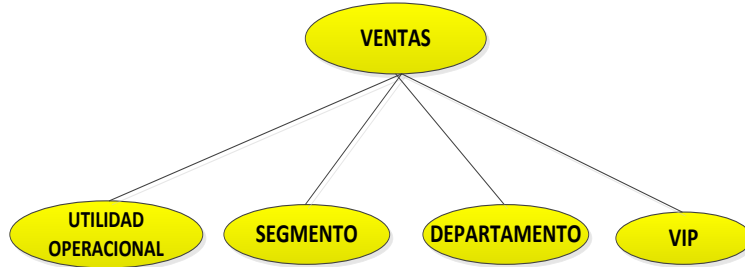
Tabla 3. Datos de los clientes para el 2011

DEPARTAMENTO	SEGMENTO	VENTAS	UTI. OPER	VIP
SANTANDER	DOTACIONES MP	ALTA	GOLD	SI
DISTRITO CAPITAL	INDUSTRIAL MP	MEDIA	GOLD	NO
SANTANDER	DOTACIONES MP	ALTA	GOLD	SI
DISTRITO CAPITAL	INDUSTRIAL MP	ALTA	GOLD	SI
SANTANDER	INDUSTRIAL MP	ALTA	GOLD	SI
DISTRITO CAPITAL	INDUSTRIAL MP	ALTA	GOLD	SI
SANTANDER	INFANTIL MP	MEDIA	PLATINUN	NO
RISARALDA	FEMENINA MP	MEDIA	PLATINUN	NO
SANTANDER	INDUSTRIAL MP	MEDIA	PLATINUN	NO
DISTRITO CAPITAL	FEMENINA MP	MEDIA	PLATINUN	NO
RISARALDA	INFANTIL MP	MEDIA	PLATINUN	NO
DISTRITO CAPITAL	INDUSTRIAL MP	MEDIA	NORMAL	NO
DISTRITO CAPITAL	DOTACIONES MP	MEDIA	PLATINUN	NO
DISTRITO CAPITAL	INDUSTRIAL MP	MEDIA	NORMAL	NO
DISTRITO CAPITAL	DOTACIONES MP	BAJA	NORMAL	NO
SANTANDER	DOTACIONES MP	BAJA	NORMAL	NO
RISARALDA	FEMENINA MP	BAJA	NORMAL	NO
SANTANDER	INFANTIL MP	BAJA	NORMAL	NO
SANTANDER	FEMENINA MP	BAJA	NORMAL	NO
DISTRITO CAPITAL	INFANTIL MP	BAJA	NORMAL	NO

Fuente. Autor

La estructura de la red bayesiana se muestra en la figura 6.

Figura 6. Red construida



Fuente. Autor

Las probabilidades condicionales para cada una de las variables son:

- Ventas

ALTA	BAJA	MEDIA
0,25	0,3	0,45

- Utilidad

	UTILIDAD		
VENTAS	GOLD	NORMAL	PLATINUN
ALTA	1	0	0
BAJA	0	1	0
MEDIA	0,11	0,22	0,67

- VIP

	VIP	
VENTAS	SI	NO
ALTA	1	0
BAJA	0	1
MEDIA	0	1

- Segmento

	SEGMENTO			
VENTAS	DOTACIONES	FEMENINA	INDUSTRIAL	INFANTIL
ALTA	0,4	0	0,6	0
BAJA	0,33	0,33	0	0,34
MEDIA	0,12	0,22	0,44	0,22

- Departamento

	DEPARTAMENTO		
VENTAS	DC	RISARALDA	SANTANDER
ALTA	0,4	0	0,6
BAJA	0,33	0,17	0,5
MEDIA	0,56	0,22	0,22

Ahora, para calcular la probabilidad condicionada de tener un cliente con ventas *Altas* dado que el departamento es *Santander*, la utilidad operacional es *Gold*, el segmento es *Dotaciones* y es un cliente *VIP* se toma el conjunto *O* como:

$$O = (Dep = Sant, UO = Gold, Seg = Dot, VIP = Si)$$

Aplicando el teorema de Bayes se tiene:

$$P(V = alta|O) = \frac{P(O|V = alta) P(V = alta)}{P(O)}$$

donde,

$$P(O) = P(O|V = alta) P(V = alta) + P(O|V = media)P(V = media) + P(O|V = baja)P(V = baja)$$

Al encontrar cada una de las probabilidades y remplazar en (1) se tiene:

$$P(V = alta|O) = \frac{0,24 * 0,25}{0,06}$$

$$P(V = alta|O) = 1$$

Es decir, tenemos la certeza de que las ventas son *altas* cuando se tiene el evento O.

3.2.1.2 Árboles de decisión.

Los árboles de decisión constituyen uno de los métodos del aprendizaje inductivo supervisado más utilizados en diversas áreas de la ciencia e ingeniería. Una de las principales virtudes, es la sencillez de los modelos obtenidos. Un árbol de decisión, en su definición más sencilla, es una estructura que visualmente describe una serie de condiciones que conducen a la toma de una decisión.

Según *Hernández Orallo et al. (2004)*, son un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas.

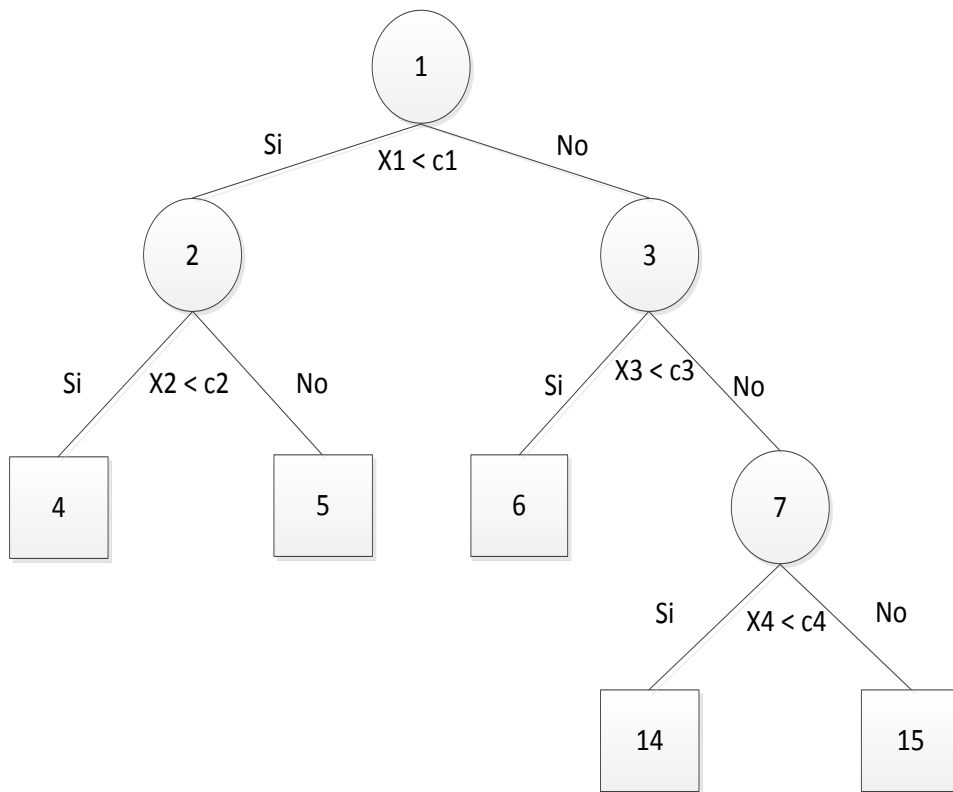
Existen 2 tipos de árboles de decisión:

- *Árboles de clasificación*: rotulan los registros y los asignan a la clase correspondiente; pueden dar una confianza de que la clasificación sea correcta. En este caso el árbol de clasificación da la probabilidad de clase, es decir, la probabilidad de que ese registro pertenezca a una clase determinada.
- *Árboles de regresión*: estiman el valor de una variable objetivo que toma valores numéricos. Dadas unas variables de entrada, el árbol estima el

resultado de la combinación de estas y asigna un valor de salida para la variable dependiente.

La tarea para la cual los árboles de decisión se adecuan mejor es para la clasificación; la estructura de condición y ramificación de un árbol de decisión es idónea para este problema. La característica más importante de los árboles de decisión es que las clases son disjuntas, es decir, un ejemplo será de la clase *a* o de la clase *b*, pero no podrá ser al mismo tiempo de las clases *a* y *b*. (Ver figura 7)

Figura 7. Esquema de un Árbol de decisiones



Fuente. Autor

Existen dos aspectos importantes en los árboles de decisión y sobre los cuales radica el buen funcionamiento de los algoritmos:

1. Particiones posibles
2. Criterio de selecciones de particiones

❖ Particiones Posibles

Las condiciones de partición son exhaustivas y excluyentes. Cuantos más tipos de condiciones se permitan, más posibilidad se tendrá de encontrar los patrones que hay detrás de los datos. A medida que el número de particiones aumente los árboles de decisión serán más expresivos, y probablemente, más precisos. Si bien en la medida que aumentan las particiones la complejidad del algoritmo también va a aumentar, el *quid* de la cuestión es encontrar una buena relación entre la expresividad y la eficiencia del árbol de decisión¹³.

Inicialmente todos los datos están juntos (nodo raíz) y el algoritmo parte los datos utilizando cada división binaria posible en cada campo. El algoritmo busca la división que fraccione los datos en dos partes que son más puras que la original. Esta división es aplicada a cada nueva caja y continúa hasta no encontrar más divisiones útiles. Las reglas de partición en un nodo dependen exclusivamente de los atributos, por lo cual, son las mismas tanto para clasificación como para regresión.

Para encontrar la división inicial se empieza con un conjunto de datos de entrenamiento consistente en registros preclasificados; esto significa que el campo objetivo o variable dependiente tiene una clase conocida. La meta es construir un árbol que distinga entre las clases. El árbol puede ser utilizado para asignar una clase al campo objetivo de nuevos registros, basado en los valores de otros campos o variables independientes.

❖ Criterio de Selección de Particiones

La mayoría de los criterios se basan en obtener medidas derivadas de las frecuencias relativas de las clases en cada uno de los hijos de la partición respecto a las frecuencias relativas de las clases en el padre. Por ejemplo, una buena partición será aquella en que se parte de un nodo que contenga 50% de

¹³HERNÁNDEZ, José et al. Introducción a la minería de datos, 2004.

la clase p y 50% de la clase n y produzca dos nodos a_1 y a_2 en donde todos los de la clase p pertenecen al nodo a_1 y todos los de la clase n pertenecen al nodo a_2 . Es una buena partición ya que estos dos nodos son más puros que el nodo padre.

La primera tarea es definir cuál de los campos independientes realiza la mejor división, la cual está definida como aquella que mejor separa los registros en grupos donde predomina una clase única. La medida utilizada para evaluar un divisor potencial es la reducción en *diversidad* (“medida de impureza” o “incremento de pureza”). Éste índice corresponde a la probabilidad de que el segundo registro pertenezca a una clase diferente de la primera y está dado por el nivel de entropía

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I(p_i, n_i)$$

Para buscar el mejor divisor en un nodo, el algoritmo del árbol de decisión considera cada campo de entrada a su vez. En esencia cada campo es clasificado, entonces cada posible decisión es ensayada y será mejor aquella que tiene mayor disminución de *diversidad*. Esto se repite para todos los campos y el ganador se escoge como el divisor del nodo.

Para encontrar la variable más informativa, el criterio seleccionado está basado en la ganancia de información; la información dada por un objeto o elemento depende de su probabilidad y puede ser medida en *bits*. Siendo C un conjunto de objetos que contiene p objetos de la clase P y n objetos de la clase N , se tiene que la cantidad de información producida por la variable dependiente es:

$$I(p, n) = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

La ganancia de información que permite obtener la variable que mejor clasifica los datos es:

$$gain(A) = I(p, n) - E(A)$$

La división inicial produce dos nodos, cada uno de los cuales es dividido de forma similar al nodo raíz. Si las salidas del nodo son iguales no tiene sentido dividirlo, en caso tal el nodo se rotula como un *nodo hoja*.

Distintas modificaciones a los árboles de decisión han dado como resultado la aplicación de éstos para diferentes tareas como, por ejemplo, regresión, agrupamiento o la estimación de probabilidades.

❖ *CART (Classification and Regression Trees)*

El método CART es un algoritmo de árbol de decisión que realiza particiones binarias sobre los atributos, pero que va asignando una media y una varianza a cada nodo, intentando seleccionar las particiones que reduzcan las varianzas de los nodos hijos. Se trata de un proceso recursivo, que se repite hasta alcanzar el criterio de homogeneidad o hasta llegar a otro criterio de detección.

La misma variable predictora puede ser utilizada varias veces en distintos niveles del árbol. CART es un método no lineal y no paramétrico que ofrece un índice de importancia a las variables.

La diferencia en la construcción de un CART a un árbol de decisión para clasificación radica en:

- La función aprendida tiene dominio real y no discreto
- Los nodos hoja del árbol se etiquetan con valores reales, de tal manera que una cierta medida de la calidad se maximice, por ejemplo la varianza de los ejemplos que caen en ese nodo respecto al valor asignado.

Entre las ventajas de los CART se tienen:

- Es un método bastante flexible.
- CART no está afectado por valores extremos, colinealidad, heterocedasticidad que afecten los procedimientos establecidos. Los valores outliers pueden ser aislados en un nodo y no tienen ningún efecto en la división.
- Puede revelar y detectar interacciones en un conjunto de datos.
- Produce resultados utilizando solo las variables más importantes.

Algunas desventajas de este tipo de árbol son:

- Al ser un algoritmo binario, tiende a generar árboles de muchos niveles.
- No está basado en un modelo probabilístico. No existen intervalos de confianza asociados con las predicciones derivadas usando el algoritmo para clasificar un conjunto de datos.

Ejemplo 2: Construcción de un árbol de decisiones

La tabla 4 representa un conjunto de datos para la clasificación de clientes de acuerdo a criterios como: *departamento*, *segmento*, *algodón* y *VIP*. La clasificación se realiza en función del atributo *VIP* que determina si un cliente es muy importante para la empresa y recibirá un trato especial. Luego de la construcción del modelo se podrá determinar qué clientes pueden ser *VIP*.

De acuerdo al atributo que se decidió para la clasificación se tienen 2 clases, la clase P (*VIP=Si*) y la clase N (*VIP=No*).

La *ganancia de información* se calcula como sigue:

- Calcular la cantidad de información de acuerdo a las clases que se forman en el atributo que se tomó para la clasificación (*VIP*).

$$I(11,9) = -\left(\frac{11}{11+9} \log_2 \frac{11}{11+9}\right) - \left(\frac{9}{11+9} \log_2 \frac{9}{11+9}\right)$$

$$I(11,9) = 0,9928$$

Tabla 4. Datos de entrenamiento para la clasificación de clientes VIP

DEPARTAMENTO	SEGMENTO	ALGODÓN	VIP
SANTANDER	DOTACIONES MP	Si	Si
DISTRITO CAPITAL	INDUSTRIAL MP	No	Si
RISARALDA	FEMENINA MP	Si	Si
SANTANDER	INFANTIL MP	Si	No
DISTRITO CAPITAL	INDUSTRIAL MP	Si	Si
SANTANDER	DOTACIONES MP	No	No
DISTRITO CAPITAL	DOTACIONES MP	No	Si
DISTRITO CAPITAL	INDUSTRIAL MP	No	No
SANTANDER	DOTACIONES MP	No	Si
DISTRITO CAPITAL	INDUSTRIAL MP	Si	Si
DISTRITO CAPITAL	INDUSTRIAL MP	Si	No
SANTANDER	FEMENINA MP	No	Si
SANTANDER	INFANTIL MP	Si	No
RISARALDA	FEMENINA MP	No	No
RISARALDA	INFANTIL MP	Si	Si
SANTANDER	INDUSTRIAL MP	Si	Si
DISTRITO CAPITAL	DOTACIONES MP	No	No
SANTANDER	INDUSTRIAL MP	No	Si
DISTRITO CAPITAL	FEMENINA MP	Si	No
DISTRITO CAPITAL	INFANTIL MP	Si	No

Fuente. Autor

- Luego la entropía para el atributo *departamento* es:

$$E(Dep) = \frac{8}{20} (0,9544) + \frac{9}{3} (0,9911) + \frac{3}{20} (0,9183)$$

$$E(Dep) = 0,9655$$

- Finalmente, se calcula la *ganancia de información*:

$$gain(Dep) = 0,9928 - 0,9655$$

$$gain(Dep) = 0,0273$$

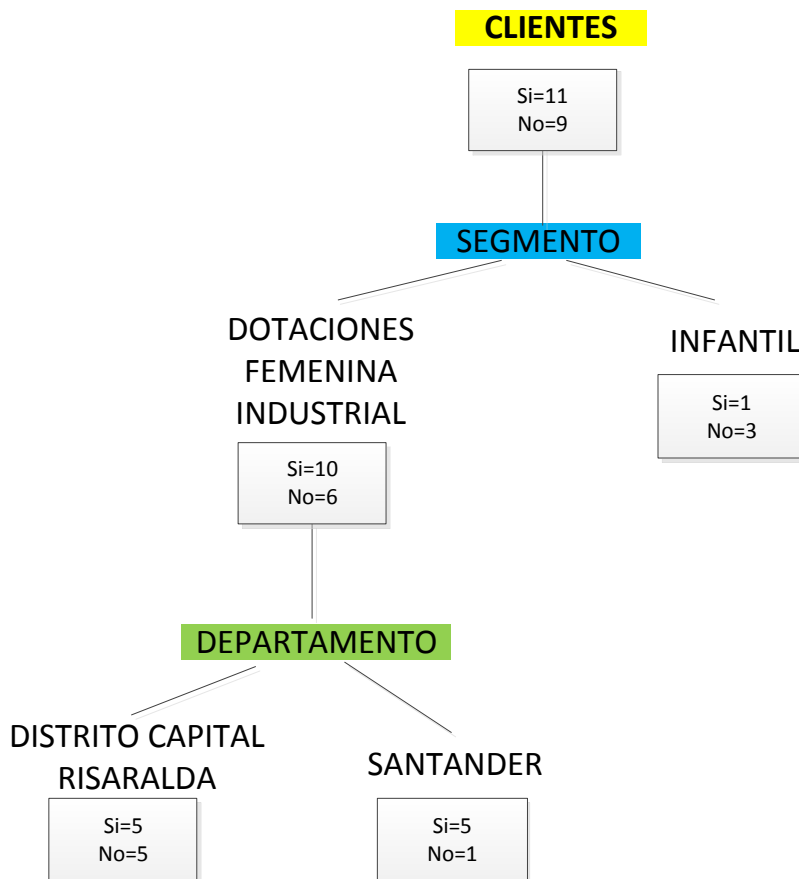
El proceso se repite para los demás atributos y se selecciona el que ofrece mayor ganancia de información. La ganancia de información para los restantes atributos es:

$$gain(Segmento) = 0,0857$$

$$gain(Algodón) = 0,00007$$

La primera partición del árbol se realiza sobre el atributo *segmento*, la asignación de los segmentos a cada uno de los dos grupos se hace basada en criterios de homogeneidad. El proceso de construcción del árbol se muestra a continuación:

Figura 8. Árbol de decisiones (parcial)



Fuente. Autor

3.2.1.3 Análisis de Clúster (Clustering).

Es una tarea de minería de datos que trata de encontrar grupos entre un conjunto de individuos u objetos a través de las características que estos poseen. El objetivo del análisis clúster es clasificar a los objetos de tal forma que cada objeto sea muy parecido a los que hay en su grupo. Los grupos resultantes deben mostrar bastante homogeneidad entre los elementos del grupo y un alto grado de heterogeneidad entre los diferentes grupos.

El análisis de clúster es un método que permite descubrir asociaciones y estructuras en los datos que no son evidentes a priori pero que pueden ser útiles una vez que se han encontrado. Los resultados pueden contribuir a la definición formal de un esquema de clasificación, a sugerir modelos estadísticos para describir poblaciones, asignar nuevos individuos a las clases para diagnóstico e identificación, etc.

Podemos encontrar dos tipos fundamentales de métodos de clasificación: **Jerárquicos** y **No Jerárquicos**. En los primeros, la clasificación resultante tiene un número creciente de clases anidadas mientras que en el segundo las clases no son anidadas. Los métodos pueden dividirse en **aglomerativos** y **divisivos**. En los primeros se parte de tantas clases como objetos tengamos que clasificar y en pasos sucesivos van obteniendo clases de objetos similares, mientras que en los segundos se parte de una única clase formada por todos los objetos que se va dividiendo en clases sucesivamente.

❖ Elección de Variables

La selección de las variables, al igual que en las técnicas anteriores, es muy importante pues sobre cualquier objeto o individuo es posible encontrar un gran número de variables, pero esto no siempre resulta útil, ya que la inclusión de variables irrelevantes puede no ser contrastada por el análisis de clúster y además aumenta la posibilidad de errores en la conclusión final.

❖ Elección de una Medida de Asociación

Para poder unir los objetos o individuos en grupos se debe seleccionar una medida de similaridad o disimilaridad, de tal manera que ésta nos marque la relación o diferencia entre los objetos o individuos. La función distancia y coeficientes de asociación se usan de manera amplia para la elección de las medidas de disimilaridad/ similaridad.

La función d se define sobre cierto conjunto M no vacío, como:

$$d: M \times M \rightarrow R$$

el par $(x, y) \rightarrow$ se asocia con el número $d(x, y)$

aquí, R es el conjunto de los números reales, la función d satisface las siguientes condiciones para tres elementos cualesquiera $x, y, z \in M$ se tiene

$$1. d(x, y) \geq 0 \text{ y } d(x, y) = 0 \text{ si y sólo si } x = y$$

$$2. d(x, y) = d(y, x) \text{ (Simetría)}$$

$$3. d(x, y) \leq d(x, z) + d(z, y) \text{ (Desigualdad triangular)}$$

La primera condición exige que la distancia entre dos sujetos debe ser mayor que cero y solo valdrá cero cuando se calcula a partir del mismo sujeto, la segunda condición dice que la distancia entre los sujetos x y y es la misma que entre los sujetos y y x ; la última indica que siempre debe ser menor o igual al del sujeto x al sujeto y que hacer el recorrido a través de z .

Los siguientes son ejemplos de funciones distancia, llamadas también métricas; definidas sobre el conjunto $R^n = \{x | x = (x_1, x_2, \dots, x_n); x_i \in R \ i = 1, \dots, n\}$

1. $d_p(x, y) = (\sum_{k=1}^n |x_k - y_k|^p)^{1/p}$ Distancia de Minkowsky
2. $d_1(x, y) = \sum_{k=1}^n |x_k - y_k|$; Distancia del taxista o de la ciudad
3. $d_2(x, y) = (\sum_{k=1}^n |x_k - y_k|^2)^{1/2}$; Distancia Euclídea
4. $d(x_i, x_j) = \sqrt{(x_i - x_j)V^{-1}(x_i - x_j)}$; Distancia de Mahalanobis: siendo V la matriz de varianza de las variables de segmentación.

Formalmente la similitud se define como una función no negativa y simétrica entre dos observaciones x_i y x_j denotada por s_{ij} que satisface las siguientes propiedades:

1. $s_{ij} = 1$
2. $0 \leq s_{ij} \leq 1$
3. $s_{ij} = s_{ji}$

La similitud es una medida adecuada para establecer el grado de similaridad entre los individuos, cuyas características han sido tomadas en escala nominal.

Cuando se elige una distancia como medida de asociación los grupos se formarán como aquellos más parecidos, es decir, que la distancia sea la mínima, generalmente es usado en aquellos datos que son medibles.

Cuando se elige una medida de similaridad los grupos se formarán maximizando la similaridad. De tal manera que existirán muchos tipos diferentes de distancias y similaridad y dependiendo de cada circunstancia se elegirán una u otra. Es usado en variables no medibles.

Existen diferentes algoritmos para la aplicación del análisis de clúster y entre ellos el más popular es el K-MEDIAS el cual parte de un número determinado

de prototipos y de un conjunto de ejemplos a agrupar, sin etiquetar. La idea es situar a los prototipos en centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares.

3.2.2 Minería de datos en los negocios.

Con el uso creciente de computadores hay un volumen creciente de datos que están siendo generados y almacenados. Esto puede llevar a las empresas a convertirse en “ricos en datos y pobres en información”¹⁴. La minería de datos ha demostrado ser muy eficaz en muchos campos, incluyendo los negocios.

El apoyo de la minería de datos se requiere para dar sentido a las masas de datos generadas en el negocio por la tecnología informática. La comprensión de este sistema de generación de información y las herramientas disponibles que conducen al análisis son fundamentales para los negocios en el siglo XXI.

Entre las oportunidades de negocio para las empresas que cuenten con una base de datos de calidad se tienen:

- Predicción automática de comportamientos: por ejemplo, la MD usa los resultados de campañas de marketing realizadas con anterioridad para identificar el perfil de los clientes que son más propensos a comprar el producto y así sustituir el correo masivo por el correo dirigido.
- Predicción automática de tendencias: basados en datos históricos, se creará un modelo para predecir las tendencias. Por ejemplo ventas en el futuro.

La MD, ha sido fundamental en la gestión de las relaciones con los clientes, en el análisis financiero y muchas otras áreas y sectores de diferentes industrias. (Ver tabla 5)

¹⁴SHORTLAND, R; SCARFE, R. Data Mining applications in BT. En : BT Technology journal. Vol. 25, Nos 3 y 4 (2007).

Tabla 5. Industrias y áreas de aplicación de la MD

Industria	Área de Aplicación
Medicina	Biomedicina, efectos secundarios de los medicamentos, la contención de costos del hospital, el análisis de la secuencia genética y la predicción
Finanzas	Aprobación del crédito, predicción de la quiebra, predicción del mercado de valores, detección de fraudes, detección de acceso no autorizado a los datos de crédito, la selección de fondos de inversión
Agricultura	Clasificación de las enfermedades del tomate y la soya
Social	Datos demográficos, tendencias de votación, resultados electorales
Mercadeo y ventas	Identificación de subgrupos socio-económicos, muestras de comportamiento inusual, patrones de compras al por menor, análisis del producto, frecuencia de los patrones, predicción de las ventas
Seguros	Detección de reclamaciones fraudulentas y excesivas, confirma la separación
Ingeniería	Diagnóstico de sistemas expertos de automóviles, Telescopio Espacial Hubble, diseño asistido por computador(CAD) de bases de datos, estimaciones de empleo
Física y Química	Electroquímica, investigación de superconductores
Militar	Análisis de inteligencia, fusión de datos y otras aplicaciones clasificadas
Derecho	Bienestar y fraude fiscal, coincidencia de huellas dactilares, recuperación de vehículos robados

Fuente. Data Mining applications in BT. Shortland R., and Scarfe R. 2007

El interés del mundo profesional por la minería de datos se manifiesta por el variado y amplio conjunto de empresas que han lanzado al mercado productos de *Data Mining*, entre las cuales se encuentran SPSS Modeler (antes Clementine) y SAS (Enterprise Miner).

3.2.3 Minería de datos y su relación con el CRM.

Cada vez se vuelve más frecuente escuchar en las organizaciones palabras como *Almacén/bodega de Datos (Data Warehousing)*, *Software de Gestión y Minería de Datos* todo bajo el marco de la gestión de las relaciones con el cliente (CRM, Customer Relationship Management) y su aplicación en beneficio de la obtención de una ventaja competitiva. Los clientes son el activo más

importante de una organización, por tanto, no puede haber ninguna perspectiva de negocio sin clientes satisfechos que permanezcan leales y desarrollen una relación perenne y de valor con la organización.

CRM es un conjunto de procesos y sistemas que permiten apoyar a una estrategia de negocios a construir relaciones de largo plazo con clientes específicos (Ling y Yen, 2001). Los datos del cliente y las tecnologías de información (TI) son la base sobre la cual se construye cualquier estrategia de CRM exitosa. Además, el rápido crecimiento de internet y sus tecnologías asociadas han aumentado considerablemente las oportunidades de comercialización y han transformado la forma en las relaciones entre las empresas y su gestión de clientes.¹⁵

El enfoque organizacional hacia el cliente como forma de aumentar el valor de éste para la organización puede realizarse desde 3 aspectos importantes:

- Aumentar el uso (compras) de los productos que ya tienen
- Vender más productos o con más margen
- Mantener los clientes por un periodo de tiempo más largo

La estructura de un CRM se define en *operacional* y *analítica*. El CRM operacional se refiere a la automatización de los procesos del negocio, mientras que el CRM analítico se refiere al análisis de las características y comportamientos de los clientes con el fin de apoyar las estrategias de la organización de la gestión de clientes. Como tal, el CRM analítico podría ayudar a una organización para discriminar mejor y más eficazmente la asignación de recursos para el grupo de clientes más rentables.

Las herramientas de minería de datos son un medio popular para analizar los datos de los clientes en el marco del CRM analítico. Con los datos de los

¹⁵E.W.T. Ngai et al. Application of data mining techniques in customer relationship management: A literature review and classification. 2005

clientes de forma global, la tecnología de minería de datos puede proporcionar inteligencia de negocios para generar nuevas oportunidades. La aplicación de herramientas de minería de datos en el CRM es una tendencia emergente en la economía mundial.

Analizar y comprender los comportamientos de los clientes y las características es la base del desarrollo de una estrategia de CRM competitivo, con el fin de adquirir y retener a los clientes potenciales y maximizar el valor del cliente. La aplicación de técnicas de minería de datos en el CRM representa una importancia significativa en una economía centrada en el cliente.

El CRM consiste en 4 dimensiones que representan un ciclo cerrado de la gestión de clientes y tienen en común el facilitar la maximización del valor del cliente a través de una relación de largo plazo con la organización. En este camino, las técnicas de MD ayudan en la extracción o detección de patrones y características ocultas en los datos que revelan el comportamiento de los clientes. Las 4 dimensiones del CRM consisten en:

- *Identificación del cliente:* incluye la población con más probabilidad de convertirse en cliente o en aquellos que son más rentables. Analiza los clientes que se están perdiendo y la forma de recuperarlos.
- *Atracción del cliente:* luego de la identificación de clientes potenciales, la empresa puede dirigir sus esfuerzos a la atracción de clientes o segmentos objetivo mediante, por ejemplo, el marketing directo.
- *Retención del cliente:* Es la preocupación central del CRM, se refiere a la comparación entre las expectativas de los clientes con su percepción de estar satisfechos y se convierte en la condición esencial para retener a los clientes. Como elementos de retención se tienen programas de fidelización y gestión de reclamaciones. El apoyo a las campañas de marketing a través de la identificación, análisis, detección y predicción de comportamiento de los clientes.

- *Desarrollo del cliente:* se refiere a la intensidad de la transacción, al valor de la transacción y la rentabilidad que genera el cliente. Los elementos de desarrollo de clientes incluyen el análisis del valor del cliente de por vida, ventas cruzadas y el análisis de la cesta de compra.

En cada una de estas dimensiones, la MD puede apoyar su gestión mediante la aplicación de diversos modelos que incluyen asociación, clasificación, agrupamiento, previsión, regresión, el descubrimiento de secuencias y la visualización.

Para tener éxito con el CRM y abordar los objetivos antes mencionados, las organizaciones necesitan ganar penetración en los clientes, conocer sus necesidades y deseos a través del análisis de datos. Aquí es donde viene el CRM analítico. El CRM analítico analiza la información del cliente para atender mejor los objetivos del CRM y entregar el mensaje adecuado al cliente adecuado. Esto implica el uso de modelos de minería de datos con el fin de evaluar el valor de los clientes, entender y predecir su comportamiento. Se trata de analizar los patrones de los datos para extraer conocimiento y optimizar las relaciones. Por ejemplo, la minería de datos puede ayudar en la retención de clientes, ya que permite la identificación oportuna de aquellos que son más valiosos con una mayor probabilidad de irse, ganando tiempo para la generación de campañas de retención específicas. Se puede apoyar el desarrollo del cliente, haciendo coincidir los productos con los clientes y una mejor orientación a la promoción de las campañas del producto. También puede ayudar a revelar los distintos segmentos de clientes, facilitando el desarrollo de nuevos productos personalizados y ofertas de productos que mejor se enfrenten a las preferencias y prioridades específicas de éstos.

Como es claro, la minería de datos representa el enlace de los datos almacenados durante muchos años a través de diversas interacciones con los clientes en diversas situaciones y la generación del conocimiento necesario

para tener éxito en los conceptos de marketing relacional con el fin de liberar el potencial de esta información.

3.3 METODOLOGÍAS PARA PROCESOS DE MINERÍA DE DATOS

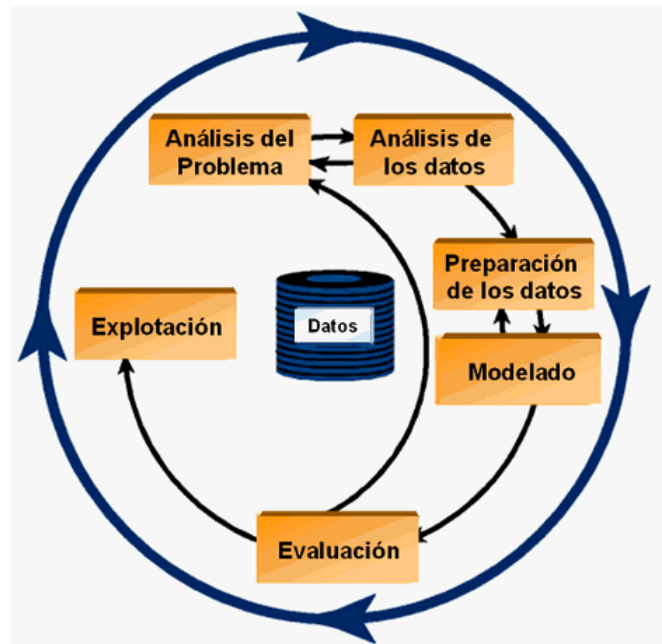
Diferentes metodologías han surgido para la aplicación de proyectos de minería de datos que pretenden facilitar la realización de nuevos proyectos, optimizar la planificación y dirección de éstos, reducir su complejidad y permitir hacer un mejor seguimiento a ellos (Gondar Nores, 2004). Entre ellas se encuentran la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*) y la metodología SEMMA (*Sample, Explore, Modify, Model, Assess*). SEMMA se centra en características técnicas del desarrollo del proceso, en tanto que CRISP-DM mantiene como foco central los objetivos empresariales del proyecto.

CRISP-DM

Está dividida en cuatro niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos. La metodología propone para un proyecto de minería de datos un ciclo de vida que consiste en seis fases. (Ver figura 9)

En el nivel más alto el proceso está organizado en 6 fases; cada fase consiste en varias tareas de nivel general. El segundo nivel se denomina *genérico* y pretende ser general para que abarque todas las situaciones posibles; las tareas generales deben ser completas y estables. En el tercer nivel se tienen tareas especializadas; describe cómo las acciones de las tareas generales se deberían desarrollar en ciertas situaciones específicas. El cuarto nivel es un conjunto de acciones, decisiones y resultados sobre el proceso de minería de datos.

Figura 9. Fases de la metodología CRISP-DM



Fuente. Manual CRISP-DM de IBM SPSS Modeler

3.4 PROBLEMAS COMBINATORIOS EN LA MINERÍA DE DATOS

Resolver un problema de optimización es encontrar la mejor solución posible a un problema formulado matemáticamente, donde el criterio que evalúa la calidad de la solución es cuantitativo, generalmente asociado a un costo y conocido como función objetivo. En los últimos años, el costo, generalmente asociado, está ligado al costo computacional el cual nos determina si un problema es fácil o no de acuerdo a los algoritmos conocidos para resolverlo.

Un algoritmo es un conjunto ordenado y finito de operaciones que permiten solucionar un problema. Un problema es fácil si existe un algoritmo que lo resuelva en tiempo polinomial, si el número de operaciones para que el algoritmo resuelva el problema es una función polinomial del tamaño del problema. Si esta función no es polinomial el problema se considera difícil. En la práctica, ha resultado muy importante por la necesidad de desarrollar estrategias para encontrar soluciones buenas a un costo computacional razonable.

En tanto los problemas se clasifican en fáciles o difíciles, los algoritmos se clasifican en exactos o completos y aproximados o heurísticos; debido a que la mayoría de los problemas de optimización combinatoria se clasifican como difíciles, la investigación ha centrado su atención en algoritmos aproximados y ha surgido la *metaheurística* que combina algoritmos heurísticos con estrategias eficientes de exploración y búsqueda.

En este sentido, las técnicas de minería de datos abordan problemas de optimización en los cuales no se hace uso de un análisis formal y riguroso del problema obteniendo soluciones de calidad a un costo razonable de recursos, principalmente tiempo.

4. PROCESO DE DESCUBRIMIENTO EN BASES DE DATOS

El presente capítulo está dedicado a la ejecución del proceso de minería de datos. Inicia con la integración y recopilación de los datos en donde se relacionan las bases de datos utilizadas y la utilidad para el proceso de caracterización de segmentos. Luego, se muestran las distintas variables utilizadas y su respectivo proceso de depuración y se finaliza con una descripción del uso de las herramientas estadísticas para el tratamiento de los datos a través de ejemplos sencillos acerca de cómo se construye un modelo en las herramientas propuestas.

4.1 INTEGRACIÓN Y RECOPIACIÓN

El proyecto de *Minería de Datos* inicia con la identificación de los posibles datos que se encuentren contenidos en diferentes recursos, ya sea en los servidores internos o producto de estudios externos como por ejemplo los demográficos, censos o investigaciones de mercado. Para este proyecto en particular se dispondrá únicamente de los datos contenidos en las bases de datos de COMERTEX S.A. junto con los datos de una investigación realizada por una empresa privada que buscaba evaluar los aspectos relacionados con el nivel de servicio así como la identificación de las razones de compra y uso de los productos. Los datos contenidos en estas y utilizados para este proyecto, corresponden a datos transaccionales del año 2.011.

Los datos son capturados y almacenados por el ERP- SEVEN y son extraídos mediante diversas plantillas previamente creadas las cuales en primera instancia son del tipo .txt y posteriormente son pasados a archivos tipo "Excel".

En este ejercicio se recopilieron las siguientes plantillas:

- Total pedidos VS Total factura (Plantilla 10)

- Consumo por producto (Plantilla 11)
- Detalles de devoluciones (PLANTILLA 38)
- Rentabilidad por cliente (PyG)
- Encuesta de satisfacción del cliente

A cada una de las plantillas mencionadas se le integraron variables que permitieran tener la mayor cantidad de información en una sola de ellas. (Ver Anexo B: Plantillas de trabajo y encuesta del nivel de servicio)

Paralelo a la integración de las bases de datos se construye el *Diccionario de Datos*, (ver depuración de bases de datos en el numeral 4.2.1), el cual contiene la definición para cada variable seleccionada.

4.2 SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN

En esta fase se analizan los datos y el tratamiento que se les debe realizar. Para este caso se trabaja únicamente con los datos de los clientes ubicados en cada uno de los 13 segmentos que actualmente maneja COMERTEX S.A.

4.2.1 Depuración de la base de datos.

Para cada una de las plantillas mencionadas anteriormente se realizó el proceso de depuración como sigue:

4.2.1.1 Total Pedidos Vs Total Factura (Plantilla 10).

Este reporte arroja la siguiente información:

- Fecha factura: es la fecha en que se genera la factura en el sistema.
- Prefijo: Indica la ciudad en donde se genera la factura.
- N° factura: Consecutivo asignado a las facturas al momento de su creación.
- Bodega: Lugar desde donde se despacha la mercancía.
- Segmento: Clasificación de clientes según su propósito comercial.

- N° pedido: Consecutivo asignado a los pedidos al momento de su creación.
- Fecha pedido: Registra el momento en que el pedido es generado.
- Fecha requerido: Si el cliente necesita la mercancía en una fecha posterior se consigna en este campo.
- Fecha liberado: El momento en que el pedido es aprobado por la gerencia financiera para ser alistado.
- Valor pedido: Indica el valor total de la mercancía solicitada por el cliente
- Valor facturado: Es el valor total de lo despachado al cliente.
- Estado pedido: En este campo se encuentra si el pedido fue cerrado, facturado o anulado.
- Nit cliente: Es el código con el cual se reconoce el cliente en la empresa.
- Ciudad cliente: Ciudad a donde se despachó la mercancía.
- Departamento cliente: Departamento a donde se despachó la mercancía.
- Código vendedor: Número que identifica al vendedor que crea el pedido.
- Dirección cliente: Ubicación exacta para el envío de la mercancía.
- Nombre cliente: Razón social registrada en el sistema de la empresa.
- Teléfono cliente: Número de contacto registrado por el vendedor.

Se procede a identificar los atributos en cada variable con el fin de eliminar datos que no hacen parte del proyecto.

❖ Fecha De Factura

Uno de los parámetros de entrada para el reporte es la fecha de la factura, por consiguiente este campo ya ha sido depurado con la información que es de interés en el proyecto, es decir, aquella que corresponde al año 2011.

❖ Prefijo

Los valores que puede tomar son los siguientes:

Tabla 6. Prefijos plantilla 10

PREFIJOS	
BTA	Bogotá
GRN	Girón
MED	Medellín
PER	Pereira
CAL	Cali
EX	Exterior
70	Girón

Fuente. COMERTEX S.A.

Las facturas con prefijo EX pertenecen a licitaciones puntuales realizadas fuera del país, las cuales no entran dentro del presente análisis, razón por la cual estos registros son eliminados.

❖ Bodega

Las bodegas que maneja el sistema se diferencian en bodegas virtuales y bodegas reales. Las bodegas virtuales se utilizan como un control de las mercancías que son entregadas en consignación, es decir, de aquellas que ya no se encuentran bajo el resguardo de COMERTEX S.A. sino que se encuentran en propiedad del cliente junto con otras bodegas virtuales que se utilizan para facturar a los proveedores con el fin de cobrar productos en mal estado.

En la tabla 7 se muestra las bodegas que registra el sistema, si son reales o virtuales y de donde se despachó realmente la mercancía. La información que se tiene en cuenta es aquella que presenta un despacho real y se muestra en la tabla 8.

Tabla 7. Bodegas que registra SEVEN-ERP

BODEGA	TIPO	DE DONDE SE DESPACHA
Almacén La 14 Calzado	Virtual	Girón
Almacén La 14 Vestuario	Virtual	Bogotá
Almacenes Bucaramanga	Real	Almacenes Propios
Almacenes Only	Virtual	Bogotá
Almaviva Espinal	Real	Almaviva Espinal
Almaviva La Caro (Almacenamiento Algodón)	Real	Almaviva La Caro (Almacenamiento Algodón)
Almaviva Medellín (Almacenamiento Algodón)	Real	Almaviva Medellín (Almacenamiento Algodón)
Bodega Carrefour Consignación	Virtual	Girón
Bodega Muestras Hogar	Virtual	No hay despacho
Bodega Muestras Vestuario	Virtual	No hay despacho
Bodega Producción Confehogar	Virtual	No hay despacho
Bodega Producción Texbymar	Virtual	No hay despacho
Bodegas Muestras Calzado	Virtual	No hay despacho
Bogotá	Real	Bogotá
Bogotá Producto No Conforme	Real	Bogotá
Bogotá Reclamación	Real	Bogotá
Comercio Exterior Expo	Virtual	Mercancía despachada para clientes de licitaciones. No será tenida en cuenta

BODEGA	TIPO	DE DONDE SE DESPACHA
Confección Natural Cotton	Virtual	No hay despacho
Creaciones Kiguz	Virtual	No hay despacho
Eventos Especiales	Virtual	Bogotá
Foltex	Virtual	No hay despacho
Girón	Real	Girón
Girón Despachos Directos	Virtual	Proveedor
Girón Producto No Conforme	Real	Girón
Insumos	Real	Insumos
Jackeline Gomez Velasquez (Pt)	Virtual	Girón
Pereira	Real	Pereira
Producción Laura Niño	Virtual	No hay despacho
Producción Producto Terminado	Virtual	No hay despacho
Sala Ventas Hogar	Virtual	No hay despacho

Fuente. COMERTEX S.A.

Tabla 8. Bodegas consideradas para el proyecto

BODEGAS A TENER EN CUENTA	
Girón	Almaviva La Caro (Almacenamiento Algodón)
Bogotá	Almaviva Medellín (Almacenamiento Algodón)
Almacenes Propios	Insumos
Almaviva Espinal	Pereira

Fuente. COMERTEX S.A.

❖ Segmento

Los valores que puede tomar esta variable son:

Tabla 9. Segmentos plantilla 10

SEGMENTOS	SEGMENTOS
Comercio Mayorista MP	Industrial MP
Comercio Minorista MP	Grandes Superficies PT
Confecciones Dotaciones MP	Mayoristas PT
Confecciones Femenina MP	Comercio PT
Confecciones Infantil MP	No Utilizar
Confecciones Masculina MP	Licitaciones
Confecciones Ropa deportiva MP	Fundaciones
Confecciones Quirúrgico MP	Empleados
Publicidad MP	

Fuente. COMERTEX S.A.

Los valores: No Utilizar, Licitaciones, Fundaciones y Empleados no se tendrán en cuenta en el estudio ya que estos clientes no son regulares.

❖ Campos Eliminados

Se eliminan los siguientes campos de la base de datos por no proporcionar mayor información acerca de los clientes:

Tabla 10. Campos eliminados plantilla 10

CAMPOS ELIMINADOS	
N. de pedido	Fecha liberado
Fecha de pedido	Valor pedido
Fecha requerido	Estado pedido
Código vendedor	Teléfono cliente

Fuente. COMERTEX S.A.

❖ Campos Agregados

Además de los datos que se obtuvieron del ERP, fue necesario crear algunos campos con el fin de complementar la información. Los campos creados son:

- Prefijo – factura: Cada prefijo tiene un consecutivo de factura diferente, es por esto que determinado número de factura se puede presentar en todos los prefijos, dando como resultado duplicados en el número de factura, por tanto, se “unen” el prefijo y la factura, para dar un único código a cada documento de factura.
- Valor bruto: Es el valor de la mercancía facturada antes de descuentos e impuestos.
- Descuento: Si al valor bruto se le aplica algún descuento especial debido al tipo de producto o al cliente al cual va dirigida la mercancía acá queda registrada la información.
- Subtotal: Es la operación matemática VALOR BRUTO – DESCUENTOS.
- IVA: De acuerdo al tipo de producto se tiene un valor de IVA distinto¹⁶.
- Costo mercancía facturada: Es el valor que se pagó al proveedor por la mercancía que se está facturando¹⁷.
- Utilidad bruta: Operación matemática SUBTOTAL – COSTO MERCANCÍA FACTURADA.
- Nit – factura: Al igual que “prefijo – factura” estos dos campos se unen con el fin de facilitar la búsqueda en las demás tablas, ya que esta combinación es única para registro.

4.2.1.2 Consumos Por Producto (Plantilla 11).

Los campos que arroja este reporte son los siguientes:

- Mes: Identifica el mes en que se realiza la venta.

¹⁶ INSUMOS: 10%, MERCANCÍA: 16%

¹⁷ Este valor es extraído del reporte CONSUMOS POR PRODUCTO (PLANTILLA 11) el cual será estudiado en el numeral 4.2.1.2.

- Código del producto: Es el código que identifica al producto dentro de la empresa, el cual es asignado automáticamente al momento de su creación en el sistema.
- Descripción del producto: Nombre asignado a cada referencia según el Manual de Codificación CM-M-02. (Ver Anexo C: Manual de codificación)
- Código bodega: Número que identifica cada una de las bodegas (reales y virtuales) que se manejan en la empresa.
- Cantidad facturada: Número de unidades despachadas al cliente del producto identificado anteriormente.
- Valor bruto: Es el valor de la mercancía facturada antes de descuentos e impuestos.
- Código área: Identificación asignada al área.
- Área: Campo que muestra el área funcional al cual está relacionado el producto.
- Número documento: Consecutivo asignado a los documentos (facturas o notas crédito) al momento de su creación.
- Tipo de documento: Identifica si el documento que se observa es una factura o una nota crédito.
- Costo unitario: Es el valor unitario que se dio al proveedor en retribución de la mercancía.
- Nit cliente: Es el código con el cual se reconoce el cliente en la empresa.
- Nombre cliente: Razón social registrada en el sistema de la empresa.
- Código vendedor: Número que identifica al vendedor que crea el pedido.
- Vendedor: Nombre del vendedor que crea el pedido.

El proceso de depuración de datos para esta plantilla es como sigue:

❖ Código de Bodega

El código de bodega identifica numéricamente la bodega usada para el almacenamiento de la mercancía. A continuación se muestran los códigos, a que bodegas pertenecen y si se incluyen dentro del estudio:

Tabla 11. Códigos de Bodega

CÓDIGO	BODEGA	SIRVE
10	Bogotá	SI
11	Bogotá Reclamación	SI
15	Almacenes Only	SI
18	Bogotá Producto No Conforme	SI
60	Pereira	SI
70	Girón	SI
75	Eventos Especiales	SI
78	Girón Producto No Conforme	SI
1100	Producción producto terminado	NO
1104	Confección Natural Cotton	NO
1134	Creaciones Kiguz	NO
1148	Foltex	NO
1151	Bodega Producción Texbymar	NO
1152	Bodega Producción Confehogar	NO

CÓDIGO	BODEGA	SIRVE
1157	Producción Laura Niño	NO
7000	Girón Despachos Directos	NO
7011	Comercio Exterior Expo	NO
7013	Insumos	SI
7025	Almacenes Bucaramanga	SI
7080	Jackeline Gómez Velasquez (Pt)	SI
7084	Almacén La 14 Vestuario	SI
7085	Bodega Carrefour Consignación	SI
7086	Almacén La 14 Calzado	SI
7087	Bodega Muestras Vestuario	NO
7088	Bodegas Muestras Calzado	NO
7089	Bodega Muestras Hogar	NO
7096	Sala Ventas Hogar	NO
8054	Almaviva La Caro (Almacenamiento Algodón)	SI
8056	Almaviva Medellín (Almacenamiento Algodón)	SI
8059	Almaviva Espinal	SI

Fuente. COMERTEX S.A.

❖ Tipo de Documento

El presente reporte arroja dos tipos de documentos: facturas y notas crédito. Las notas crédito se presentan cuando un cliente realiza una devolución o un reclamo y la mercancía se descarga de la deuda del cliente por medio de la nota crédito. Para esta variable se analizará la facturación de los clientes y para las devoluciones se trabajará con otro reporte (detalles de devoluciones Plantilla 38). En el presente sólo interesa la facturación, por lo cual, se eliminan las notas crédito.

❖ Campos Eliminados

Se eliminan los siguientes campos de la base de datos por no proporcionar mayor información acerca de los clientes:

- Código área.
- Área.
- Tipo de documento (este se elimina solamente después de haber eliminado las notas crédito).
- Código vendedor.
- Vendedor.

❖ Campos Agregados

Además de los datos que se obtuvieron del ERP, fue necesario crear algunos campos con el fin de complementar la información. Los campos creados son:

- Línea: Los productos se clasifican en diferentes líneas. En esta variable se muestra a cual pertenece cada uno. (ver Anexo D: Definición de líneas de producto)
- Grupo: Las diferentes líneas se dividen en grupos de productos similares; este campo indica a qué grupo pertenece cada referencia. (Ver Anexo E: Clasificación de productos por grupo)

- Nit – factura: Estos dos campos se “unen” con el fin de facilitar la búsqueda en las demás tablas, ya que esta combinación es única para cada registro.

4.2.1.3 Devoluciones por referencia de producto (Plantilla 38).

Esta plantilla contiene información acerca de las devoluciones que se presentaron en el año 2011 y contiene las siguientes variables:

- Fecha: Fecha en que se realizó la devolución.
- Número de la devolución: Consecutivo de la devolución.
- Factura: Número de factura a la que corresponde la devolución.
- Código: Código de clasificación del producto.
- Referencia: Nombre del producto.
- Cantidad: Unidades que se devuelven.
- Valor unitario: Valor de cada unidad que se devuelve.
- Número de la causal: Número asignado a las causales de devolución.
- Causal de devolución: Criterio de devolución.

❖ Campos Agregados

Junto a estos campos, se incluyeron variables que relacionan cada devolución al cliente que la generó e información de ese cliente como:

- Grupo: Grupo al que pertenece el producto que se devuelve.
- Línea: Línea a la cual pertenece el producto que se devuelve.
- Valor total: El producto de *valor unitario por la cantidad*.
- NIT: Número de identificación del cliente.
- Nombre.
- Segmento.
- Observación: Observaciones y/o comentarios sobre la devolución.

Al igual que en las plantillas anteriores, para incluir variables fue necesario utilizar diferentes funciones de MS EXCEL, por ejemplo *concatenar* variables de la plantilla 10 y llevar a la plantilla 38 mediante la función *buscarv* o *consultav*.

❖ Campos eliminados

La variable tipo de transacción fue eliminada luego de la depuración en donde sólo quedó el tipo de transacción 1201 correspondiente a notas crédito por devolución.

4.2.1.4 Rentabilidad por cliente (PyG).

La información que se encuentra en este reporte es la siguiente:

- NIT: Número de identificación del cliente.
- Nombre: Nombre del cliente.
- Despachos totales: Número de despachos en el año 2011.
- Ventas: Valor total de las ventas para el año 2011.
- Descuentos: Valor total de los descuentos para el año.
- Utilidad bruta: Operación matemática *Ventas-Descuentos*.
- Margen Bruto: Operación matemática *utilidad bruta/ventas*.
- Días cartera: Promedio de días en que el cliente canceló las facturas.
- Días comisión: Días para el pago de la comisión al vendedor.
- Comisión: Valor de la comisión pagada al vendedor.
- Utilidad Después de comisión: Operación matemática *utilidad bruta – comisión*.
- Costo cartera: Costo asociado al manejo del crédito del cliente.
- Intereses cartera: Valor de los intereses cobrados al cliente.
- Utilidad después de cartera: Operación matemática *utilidad después de comisión-costo cartera+ intereses cartera*.
- Costos transporte: Costos asociados al transporte de la mercancía.
- Total costos logísticos: Costos de la operación.

- Utilidad operativa: Operación matemática *utilidad después de cartera- costos transporte- total costos logísticos*.

❖ Campos Agregados

Junto a las variables mencionadas anteriormente se incluyeron otras como:

- Departamento
- Ciudad
- Segmento
- Algodón
- VIP
- Vendedor
- Valor de las ventas por cada línea de producto

Luego de adicionar éstas variables fue necesario realizar la depuración de acuerdo a los criterios que se trabajaron en las anteriores bases de datos, por ejemplo, segmentos, departamentos, etc. se cruzaron para que en las bases de datos se manejara la misma información. Adicional, fue necesario crear algunas variables temporales que permitieran el cruce de información entre las bases de datos y otras que facilitarían el manejo estadístico como numerar cada uno de los segmentos.

4.2.1.5 Encuesta de satisfacción del cliente

Esta plantilla contiene datos relacionados con el nivel de satisfacción de los clientes, el nivel de servicio que ofrece COMERTEX S.A y evalúa diferentes criterios asociados a cada uno de los siguientes aspectos:

- Asesor Comercial
- Producto
- Entregas
- Atención Telefónica
- Ferias o Eventos

Adicionalmente, la encuesta incluye preguntas acerca de los tiempos de entrega, medios o canales de información acerca del sector textil así como las razones de compra y uso de los productos textiles.

Para este trabajo, los datos que se tienen en cuenta son aquellos relacionados directamente a cada uno de los 13 segmentos que atiende la empresa.

4.3 MINERÍA DE DATOS

En esta fase se pretende generar nuevo conocimiento acerca de los clientes de COMERTEX S.A. Las técnicas a utilizar, previamente definidas, se enmarcan dentro de las necesidades de la empresa y están dadas por la clasificación de los clientes, la obtención de relaciones entre las variables y la construcción de grupos con comportamiento homogéneo.

El uso de herramientas de minería de datos, como SPSS Modeler, ha permitido la optimización del tiempo para el procesamiento de los datos y la obtención de nueva información, en algunos casos desconocida para la empresa.

4.3.1 Redes Bayesianas en SPSS Modeler

El proceso de construcción del modelo se describe a continuación.

El primer paso para la construcción de la ruta (.str) consiste en la definición de los nodos sobre el lienzo. En la pestaña *orígenes* de la barra inferior de herramientas se define el formato en el que se encuentra el archivo que contiene los datos de estudio, éste puede estar en diversos formatos como .sav (SPSS) o .xlsx (Excel) y otros. (Ver figura 10)

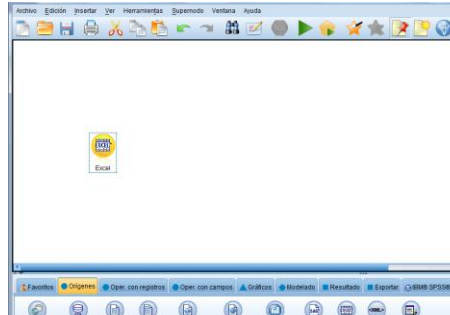
Figura 10. Paleta de nodos de origen SPSS Modeler



Fuente. SPSS Modeler

Luego de seleccionar el nodo de acuerdo tipo de archivo que se va a cargar, el nodo seleccionado es arrastrado hasta el lienzo. (Ver figura 11)

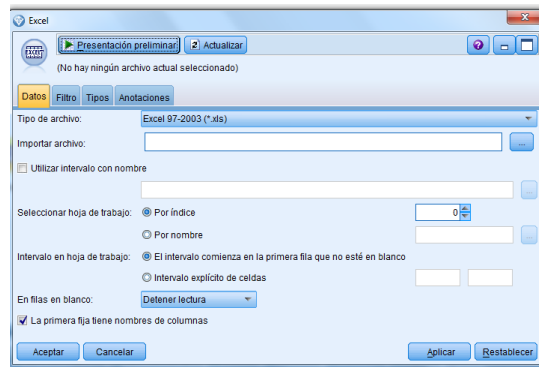
Figura 11. Nodo origen de datos sobre el lienzo



Fuente. SPSS Modeler

Para cargar la base de datos, se hace doble clic sobre el nodo del lienzo. A continuación despliega un nuevo cuadro. (Ver figura 12)

Figura 12. Datos de entrada nodo origen



Fuente. SPSS Modeler

Luego, se busca el archivo con la base de datos y se definen los aspectos que allí se requieren. En la pestaña *Tipos* se define el papel para la variable objetivo o de salida.

Luego de esto, se hace la selección de un nodo *Tipo* en el cual se define el papel de cada una de las variables para el modelo de red bayesiana a través de la pestaña *operaciones con campos* contenida en la barra inferior de herramientas. (Ver figura 13)

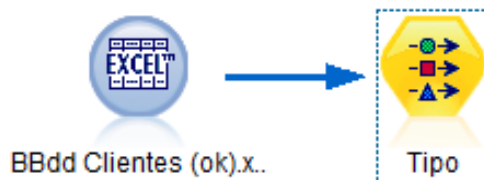
Figura 13. Nodo tipo



Fuente. SPSS Modeler

Al igual que con el nodo *origen*, el nodo *Tipo* es arrastrado hasta el lienzo y conectado con el nodo *origen*. (Ver figura 14)

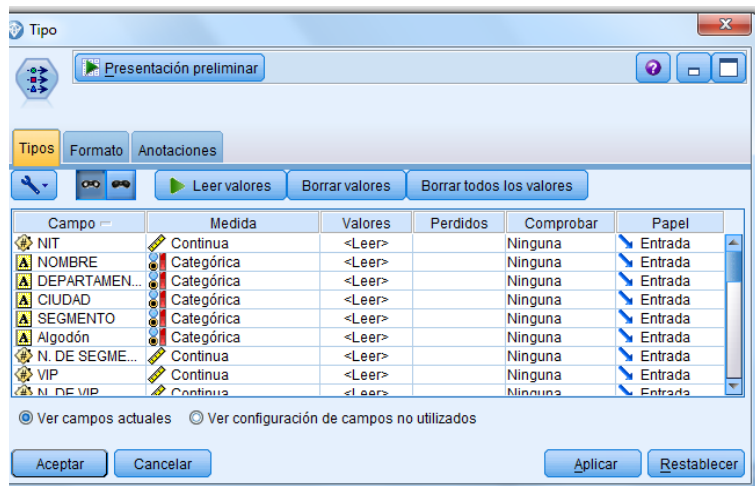
Figura 14. Conexión nodos origen-tipo



Fuente. SPSS Modeler

Para personalizar el nodo *Tipo* y definir el papel de las variables, se da doble clic sobre el nodo y éste despliega un cuadro para la personalización. (Ver figura 15)

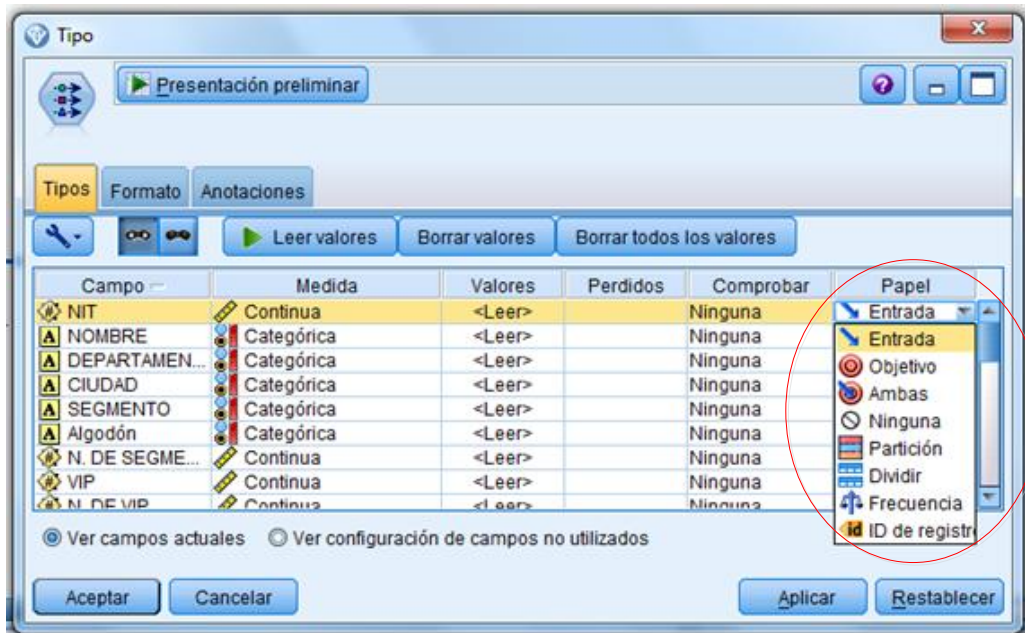
Figura 15. Personalización nodo tipo



Fuente. SPSS Modeler

Aquí se definen las variables que harán parte de la red y el papel que desempeña cada variable (Objetivo= variable de salida o respuesta; Entrada= variables que serán las predictoras del modelo). (Ver figura 16)

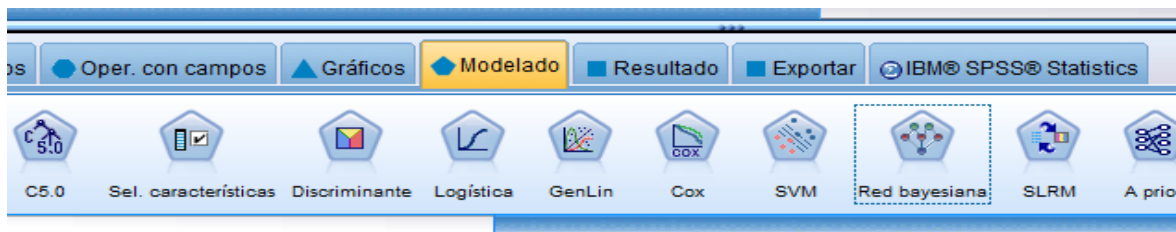
Figura 16. Selección del papel de cada variable



Fuente. SPSS Modeler

Una vez definido el papel de cada variable, ingresamos al lienzo un nodo de *Red Bayesiana* mediante el la pestaña *Modelado* de la barra inferior de herramientas. (Ver figura 17)

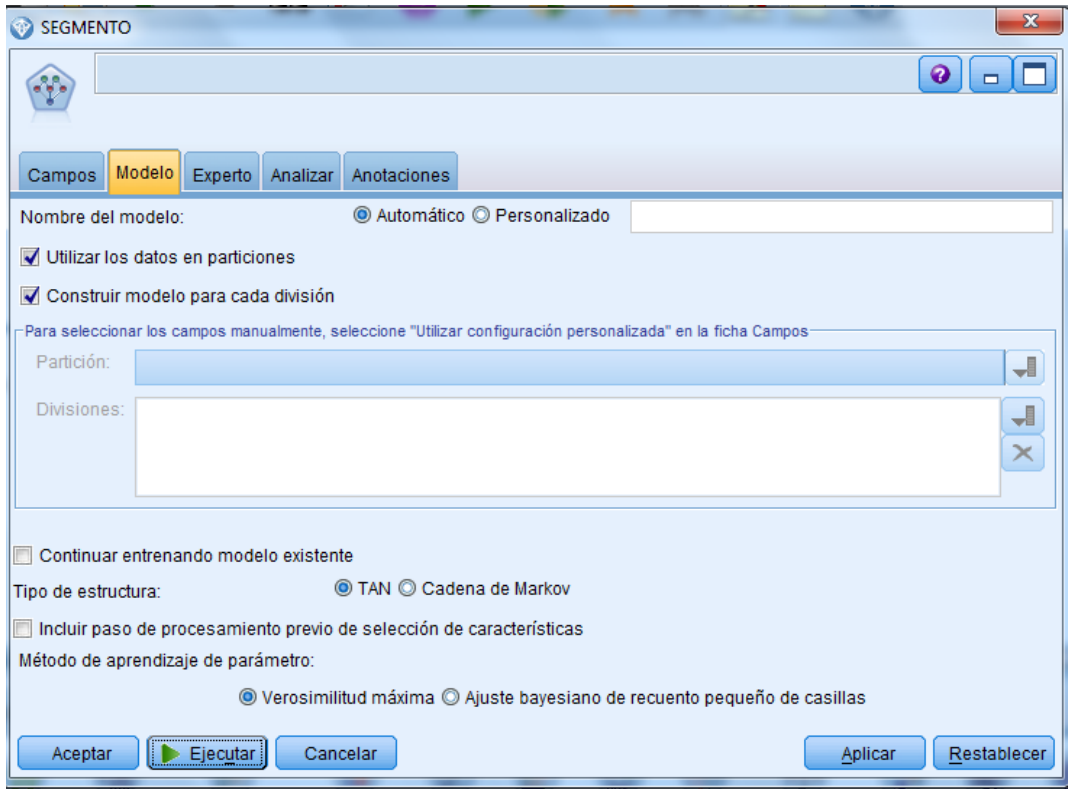
Figura 17. Nodo modelado de red bayesiana



Fuente1: SPSS Modeler

Al igual que los nodos anteriores, se lleva este al lienzo y se conecta al nodo *Tipo*. Para definir los parámetros de la red se hace doble clic sobre el nodo *red bayesiana* del lienzo y aparecerá un nuevo cuadro. (Ver figura 18)

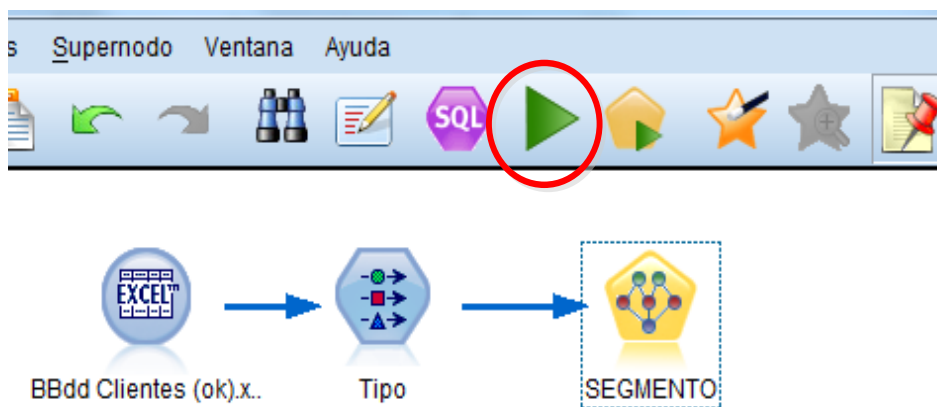
Figura 18. Parámetros para el nodo red bayesiana



Fuente. SPSS Modeler

Una vez definidos cada uno de los criterios para cada nodo, se ejecuta la ruta a través de la opción ejecutar ruta actual. (Ver figura 19)

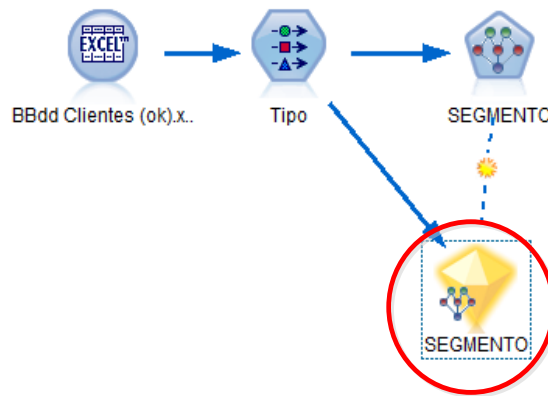
Figura 19. Ejecución de la ruta



Fuente. SPSS Modeler

Luego de la ejecución de la ruta el software produce un nugget. (Ver figura 20)

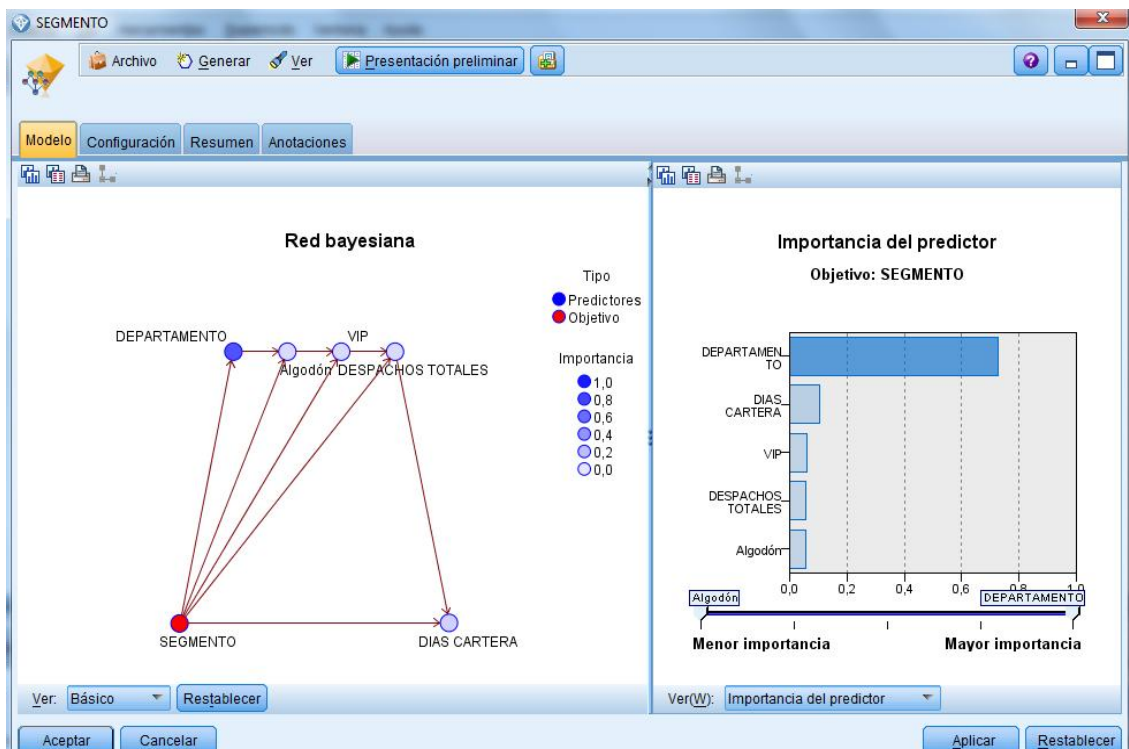
Figura 20. Nugget resultado de la ejecución de la ruta



Fuente. SPSS Modeler

Al dar doble clic sobre el nugget el software despliega un nuevo cuadro en donde muestra la red bayesiana que construyó y las probabilidades para cada variable (nodo de la red bayesiana). (Ver figura 21)

Figura 21. Resultados del modelo de red bayesiana



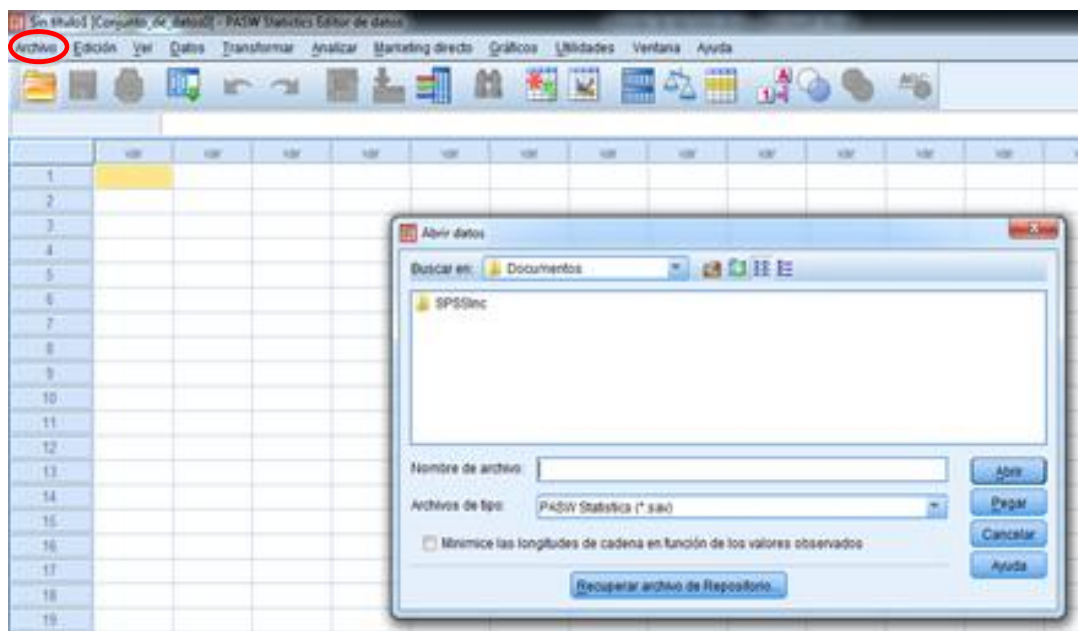
Fuente. SPSS Modeler

4.3.2 Árboles de decisión en SPSS.

La construcción de árboles de decisión mediante el software SPSS PASW o SPSS MODELER tiene el proceso que se describe a continuación. En el caso de SPSS Modeler el proceso es el mismo que en las redes bayesianas con diferencia en el nodo *Modelado* en el cual se selecciona la técnica para la construcción del modelo *CART*. A continuación se describe el proceso en SPSS PASW.

Luego de definir las variables y atributos necesarios para el estudio y la posterior depuración de los datos se procede a *importar* la base de datos desde el formato *Excel* (.xlsx) al software SPSS PASW. Siguiendo la ruta *Archivo, Abrir, Datos* y seleccionar el archivo de datos sobre el cual se construirá el árbol de decisión. (Ver figura 22)

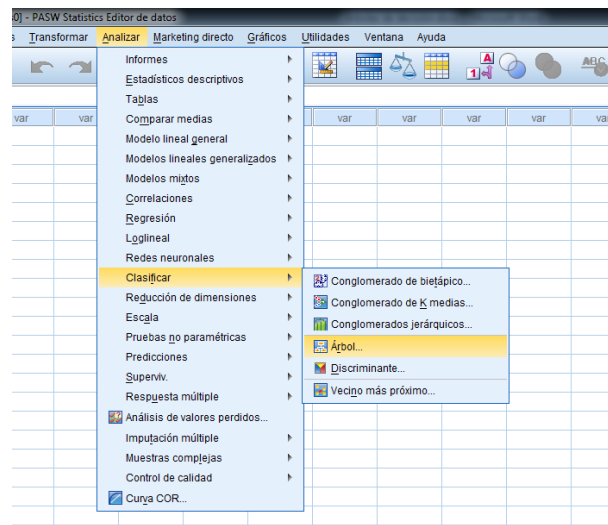
Figura 22. Ruta para cargar base de datos a SPSS PASW



Fuente. SPSS PASW

Luego de tener los datos cargados en SPSS PASW seguimos la ruta *Analizar-Clasificar-Árbol*. (Ver figura 23)

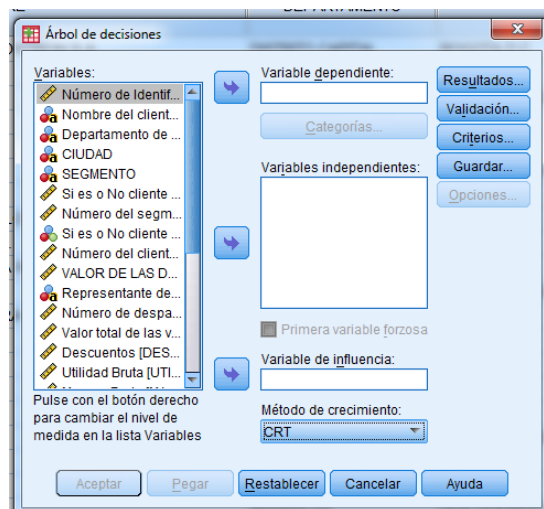
Figura 23. Ruta para crear el árbol de decisiones



Fuente. SPSS PASW

Realizado esto, se abrirá la siguiente ventana, donde se seleccionarán las variables que se incluirán en el árbol de decisiones. (Ver figura 24)

Figura 24. Definir variables del modelo



Fuente. SPSS PASW

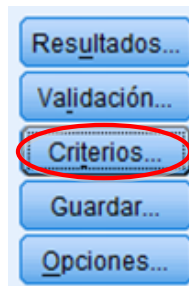
En el campo de *variable dependiente* se introduce la variable de salida del modelo, es decir, la variable que definimos como nodo raíz o nodo padre.

En el campo de *variables independientes* se introducirán aquellas variables que influirán sobre la variable dependiente.

El campo *variable de influencia* se utiliza si se quiere que la creación de nodos esté influenciada por algún criterio. En el *Método de crecimiento* definimos CRT que hace alusión a los árboles de clasificación y regresión (CART).

En las opciones, seleccionamos *Criterios*. (Ver figura 25)

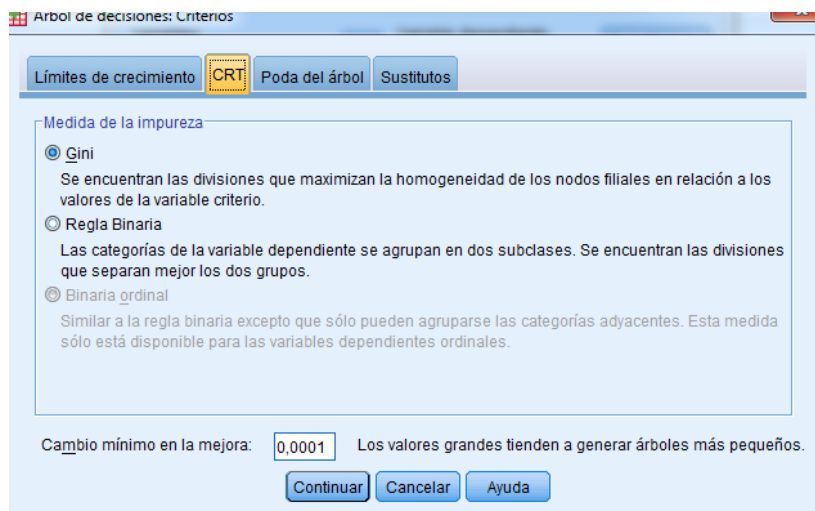
Figura 25. Ventana para definir criterios del árbol



Fuente. SPSS PASW

Luego de este paso, se abre otra ventana, en ésta se selecciona CRT y se define la medida de impureza. (Ver figura 26)

Figura 26. Criterios de definición del árbol



Fuente. SPSS PASW

Para definir el árbol que mejor clasifica los clientes de COMERTEX S.A. bajo algún criterio, definido con anterioridad, se realizan iteraciones buscando el que mejor se adapte al criterio definido.

La selección de las variables que se tomarán en cuenta para el árbol se hace mediante la elección de variables categóricas (Segmento-Línea de producto), variables escala (Ventas, devoluciones) o la combinación de estos tipos de variables. En este paso es importante definir la variable dependiente (variable del nodo raíz) y las variables independientes. También definimos el método de crecimiento del árbol que para este caso será CRT. En esta parte, es muy importante tener en cuenta que si la variable de salida es de tipo escala (continua) el árbol que se construye es de regresión, mientras que si la variable de salida es discreta el árbol será de clasificación.

De acuerdo a las variables que se ingresen, el árbol seleccionará aquella variable independiente que mejor *ganancia de información* produzca, el proceso se realiza para cada variable independiente. Una vez determinada la primera variable que dividirá el árbol, el proceso será iterativo hasta que se encuentre la mejor partición.

4.3.3 Análisis de clúster.

El análisis de clúster puede ser generado por medio de las dos herramientas descritas en las técnicas anteriores y únicamente varían en la parte de modelado. Por ejemplo, la construcción de éste análisis en SPSS Modeler se realiza de forma similar a las redes bayesianas excepto porque varía el nodo de modelado en el cual se selecciona *K-Medias*; en tanto en SPSS PASW la variación se encuentra en la ruta que sigue para encontrar la técnica de modelado que sería *Analizar-Clasificar- Conglomerado K-Medias*. De forma similar, se seleccionan las variables interés para la construcción del modelo y se definen las propiedades para cada una de ellas.

Luego de ejecutar el modelo se crearán distintos grupos dentro del conjunto de datos inicial los cuales se caracterizan porque los datos al interior de cada grupo (conglomerado) son similares en tanto, entre los distintos grupos los datos se caracterizan por una alta heterogeneidad.

5. RESULTADOS

Los resultados del presente trabajo de grado se enmarcan desde diferentes aspectos relacionados con el conocimiento del cliente, el levantamiento de información para la toma de decisiones y el deber ser de la empresa en cuanto a captura de información que optimice los procesos de toma de decisiones.

5.1 MINERÍA DE DATOS Y LEVANTAMIENTO DE INFORMACIÓN

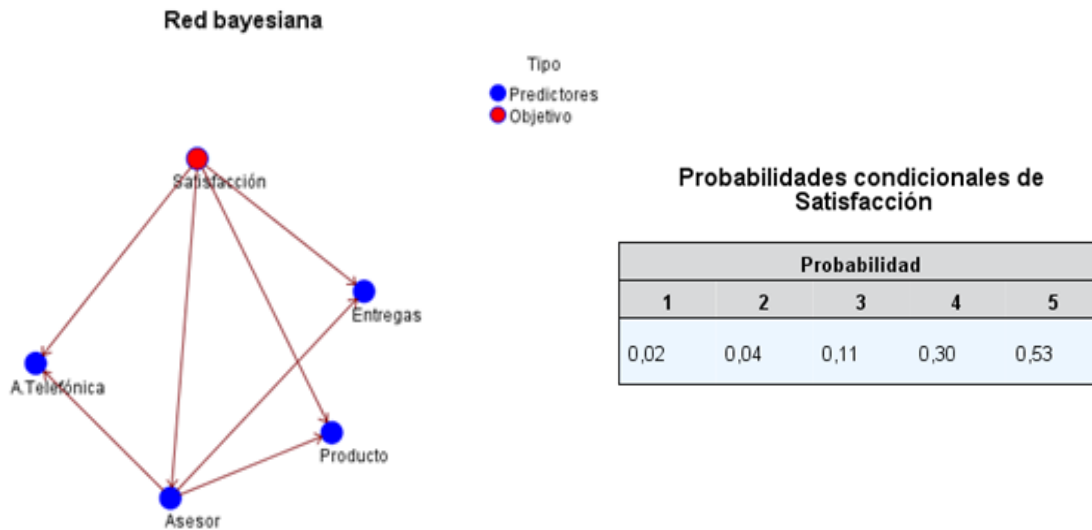
El cruce de variables y los cubos OLAP para consolidar información así como diferentes iteraciones en la aplicación de las redes bayesianas, árboles de decisión y análisis de clúster permitieron la construcción del perfil de cada segmento de mercado que actualmente atiende COMERTEX S.A.

En el caso de las redes bayesianas se construyeron desde redes sencillas que buscaban comprobar algunas dependencias/independencias entre variables específicas hasta redes complejas que buscaban la obtención de nueva información para optimizar la toma de decisiones. Algunas iteraciones se presentan a continuación como parte del proceso de construcción del perfil del segmento.

Para cada una de las redes construidas se identifica o define la variable de salida o de respuesta así como cada una de las variables predictoras. En el caso de la figura 27, el modelo se construyó a partir de la estimación del nivel de *Satisfacción* (variable de salida) y soportado en las variables predictoras *Asesor*, *Producto*, *Entrega*, *A. Telefónica*.

Este ejemplo utilizó los datos del nivel de servicio para estimar la satisfacción de los clientes basados en la calificación que ellos dieron al servicio que COMERTEX S.A. les ofrece, así mismo se identificaron las relaciones entre los aspectos evaluados.

Figura 27. Red bayesiana para satisfacción del cliente



Fuente. SPSS Modeler

El modelo obtenido da información acerca de los eventos que se pueden presentar, por ejemplo, se puede predecir el nivel de satisfacción de un cliente dada una serie de condiciones en la evaluación de los aspectos relacionados con el servicio y evidenciar la relación entre estos aspectos.

El nivel de satisfacción se evalúa de 1 a 5 en donde:

Tabla 12: Valores del nivel de satisfacción

VALOR	SATISFACCIÓN
1	Muy insatisfecho
2	Insatisfecho
3	Ni satisfecho ni insatisfecho
4	Satisfecho
5	Muy satisfecho

Fuente. COMERTEX S.A.

En la tabla 13 se observan las probabilidades de que un cliente esté satisfecho o no dada la evaluación que éste le dio al asesor comercial.

Tabla 13. Probabilidades condicionales del Asesor Comercial

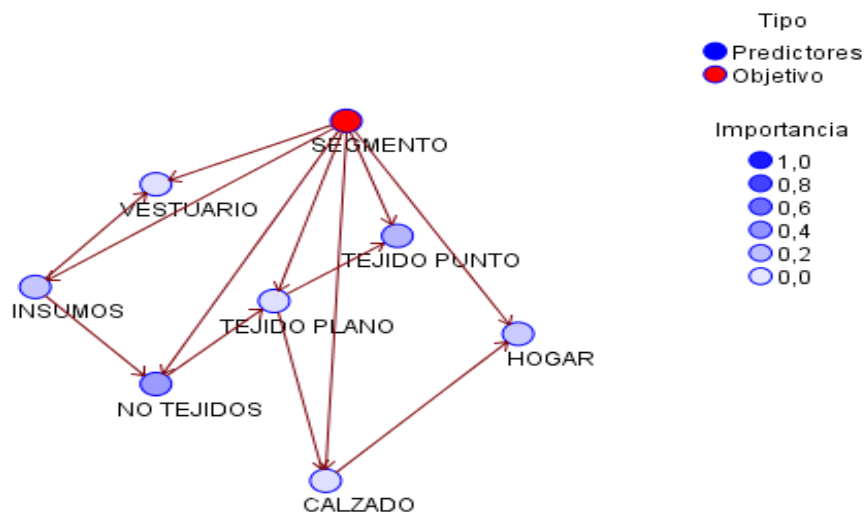
Principales	Probabilidad				
	Satisfacción < 1,8	1,8 ~ 2,6	2,6 ~ 3,4	3,4 ~ 4,2	> 4,2
1	0,29	0,00	0,00	0,57	0,14
2	0,09	0,18	0,09	0,18	0,45
3	0,00	0,00	0,06	0,47	0,47
4	0,00	0,00	0,01	0,18	0,81
5	0,00	0,01	0,01	0,09	0,89

Fuente. SPSS Modeler

Por ejemplo, si el cliente evaluó, en promedio, cada uno de los criterios asociados al asesor comercial en un rango entre 3,4 y 4,2 la probabilidad de que el cliente este muy satisfecho con el servicio es del 9%, (0,09). Así mismo, se pueden elaborar eventos que combinen los aspectos relacionados con el servicio y mediante las redes bayesianas determinar una probabilidad acerca del nivel de servicio.

Una red más compleja, como la de la figura 28, muestra la relación de independencia/dependencia entre las líneas de producto que compran los clientes de determinado segmento.

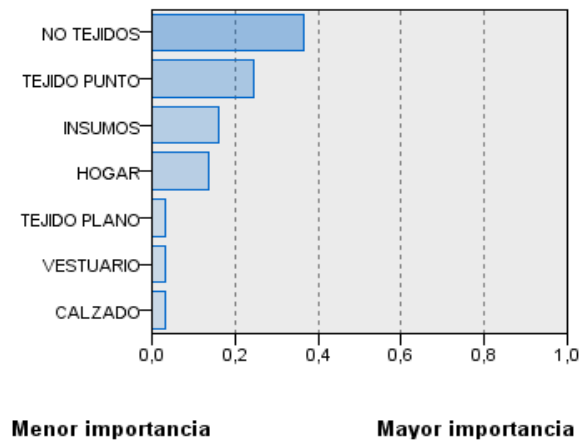
Figura 28. Red bayesiana de líneas de producto



Fuente. SPSS Modeler

La importancia de cada uno de las variables predictoras de la red se muestra en la figura 29.

Figura 29. Importancia del predictor



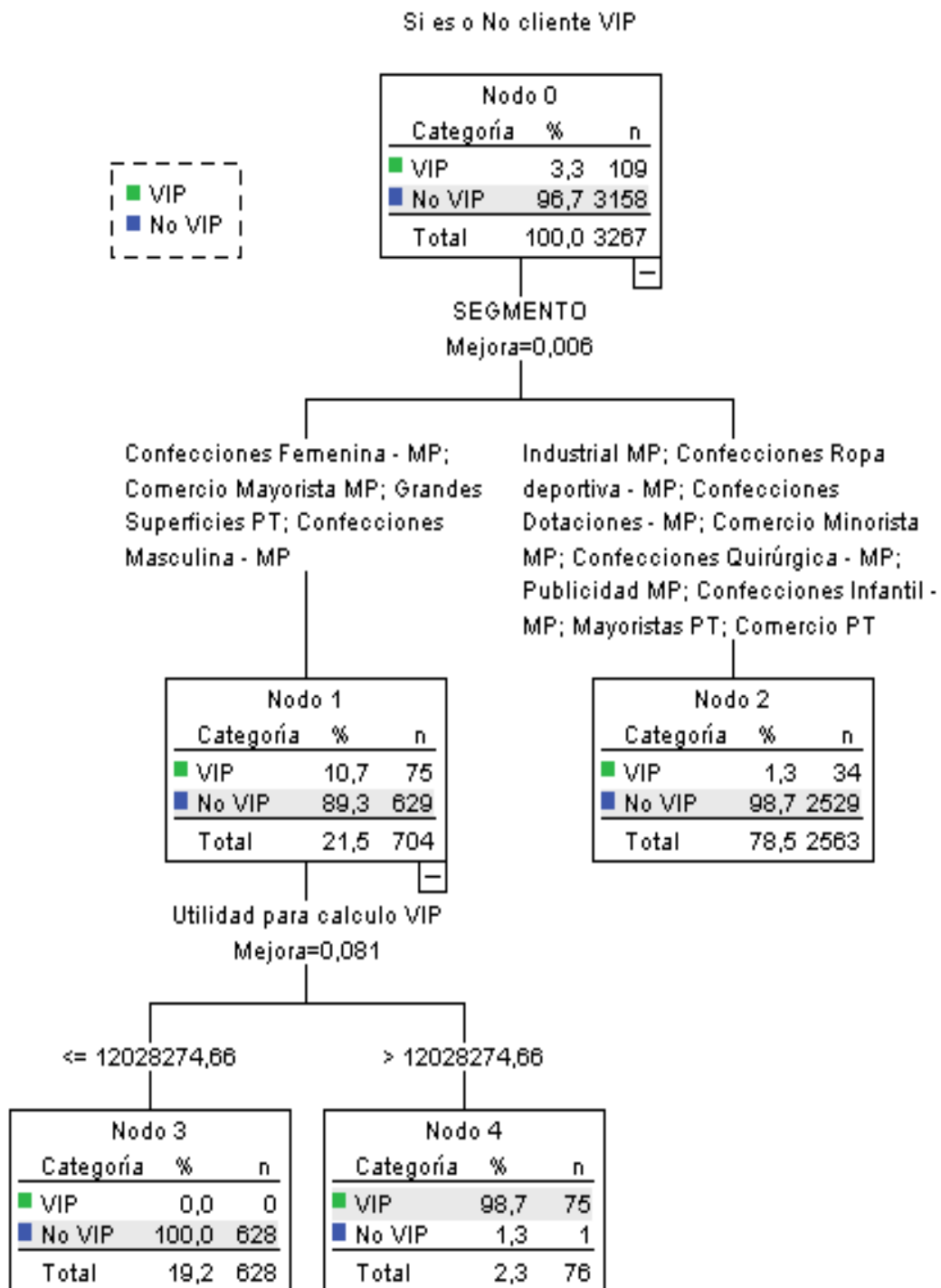
Fuente. SPSS Modeler

En el caso de esta red ofrece información acerca del comportamiento de cada línea de producto respecto al segmento, es decir, dadas las probabilidades de compra de un cliente X por línea de producto se puede inferir a que segmento Y pertenece el cliente; también se observa la relación entre las diferentes líneas de producto que facilitaran el desarrollo del proceso de venta cruzada y realizar el “análisis de la canasta” de compra para cada segmento. Diferentes iteraciones se observan en el Anexo F: Modelos de redes bayesianas.

La utilidad de los modelos de *Redes Bayesianas* está dada por la identificación de relaciones entre los aspectos relacionados con el servicio así como la importancia de cada uno de esos aspectos para la predicción del nivel general de satisfacción de los clientes y la cuantificación mediante probabilidades de las relaciones entre las líneas de producto que compra cada segmento permitiendo a la organización priorizar sobre ellos o sobre las relaciones que existen y a partir de éstos construir la promesa de servicio.

En cuanto a los árboles de decisión (CART), éstos permitieron re-clasificar los clientes a través de diferentes criterios. A continuación se muestra un árbol que los clasificó de acuerdo al tipo de cliente VIP y otras variables. (Ver figura 30)

Figura 30. CART



Fuente. SPSS Modeler

El árbol clasificó a los clientes mediante la variable *VIP* la cual da información acerca del tipo de cliente, independiente del segmento. La clasificación se realiza mediante las variables predictoras (clasificadoras) *Segmento y Utilidad Operacional*; ésta inicia de acuerdo al segmento en el que se encuentran los clientes *VIP* y posteriormente los clasifica de acuerdo al valor de la utilidad operacional para el año 2.011.

La información de este árbol se puede obtener siguiendo cada una de las ramas desde el nodo origen (*VIP*) hasta cualquier rama del árbol, por ejemplo, la clasificación muestra que el 98,7% de los clientes *VIP* generaron una utilidad operativa individual superior a \$12 millones y están ubicados en los segmentos *Confecciones Femenina y Masculina, Comercio Mayorista MP y Grandes Superficies*.

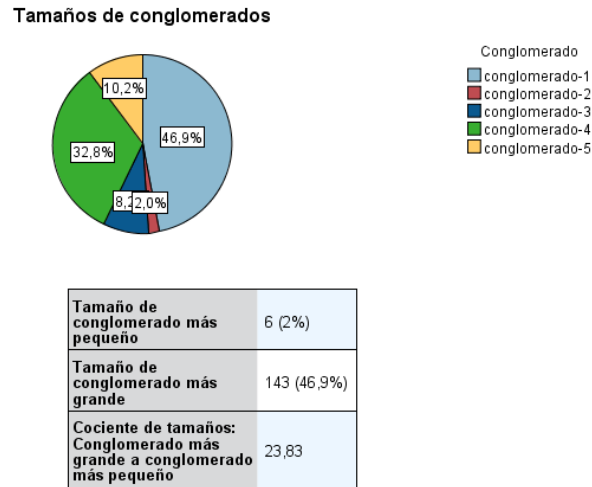
La utilidad de estas técnicas en la construcción del modelo conceptual está en la identificación de aquellos clientes con características similares que permitan clasificarlos para personalizar servicios y ofertas acordes a sus necesidades.

Al igual que en el caso de las redes bayesianas, se realizaron diferentes iteraciones buscando la mejor clasificación de los clientes. (Ver Anexo G: Clasificación de clientes mediante árboles de decisión CART).

Por otra parte, el análisis de clúster permitió agrupar los clientes en diferentes grupos con el objetivo de analizar su comportamiento. Se realizaron iteraciones buscando el número de clúster óptimo para COMERTEX S.A., por ejemplo, evaluando aspectos relacionados con el servicio como asesor comercial, producto, entregas, atención telefónica y se presenta el siguiente resultado. (Ver figuras 31 y 32)

Se construyeron 5 conglomerados que tienen una participación como se muestra en la figura 31.

Figura 31. Tamaño de los conglomerados



Fuente. SPSS Modeler

Para cada conglomerado se calculó el promedio asignado a cada aspecto evaluado como se muestra en la figura 32.

Figura 32. Importancia de los predictores para la construcción de conglomerados

Conglomerados

Importancia de entrada (predictor)
■ 1,0 ■ 0,8 ■ 0,6 ■ 0,4 ■ 0,2 ■ 0,0

Conglomerado	conglomerado-1	conglomerado-4	conglomerado-5	conglomerado-3	conglomerado-2
Etiqueta					
Descripción					
Tamaño	46,9% (143)	32,8% (100)	10,2% (31)	8,2% (25)	2,0% (6)
Entradas	A.Telefónica 4,8514	A.Telefónica 4,3492	A.Telefónica 3,1468	A.Telefónica 2,9688	A.Telefónica 2,0833
	Entregas 4,8388	Entregas 4,3385	Entregas 4,6	Entregas 3,378	Entregas 3,1
	Asesor 4,8538	Asesor 4,4916	Asesor 4,4290	Asesor 3,6820	Asesor 2,0476
	Producto 4,7317	Producto 4,0830	Producto 4,0710	Producto 3,846	Producto 2,2333

Fuente. SPSS Modeler

Los conglomerados permiten oportunidades de mejora específicas ya que se identifican las falencias en el servicio para cada segmento.

El uso de esta técnica estuvo limitado a la elaboración de una metodología para la construcción de segmentos teniendo en cuenta los comportamientos de los clientes ya que la empresa desea continuar con la segmentación actual.

A las técnicas estadísticas mencionadas se sumó el uso de estadística básica como el cruce de variables y los cubos OLAP que permitió conocer el comportamiento de los segmentos. La caracterización de cada segmento se realizó basados en la información transaccional de la empresa para el año 2011 y sobre el estudio de evaluación de satisfacción de clientes y definió aspectos como:

- Participación del segmento en cuanto a cantidad de clientes.
- Participación del segmento en cuanto a ventas.
- Ubicación de los clientes del segmento por departamentos bajo criterios como ventas o número de clientes.
- Tipos de productos que compran en cuanto a *líneas de producto* y *grupos de producto*.
- Participación de los representantes de ventas.
- Comportamiento de las ventas mes a mes.
- Grupos de productos más representativos mes a mes.
- Características de las devoluciones y su participación en el total de las devoluciones.
- Cantidad promedio de días que los clientes se toman para el pago de las facturas.
- Forma en que realizan un pedido, frecuencia y valor del pedido.
- Participación en cuanto a la utilidad operativa, cantidad de clientes que aportan utilidad positiva y negativa.
- Nivel general de satisfacción.
- Importancia de aspectos relacionados con el servicio.

- Aspectos importantes a la hora de la compra.
- Razones de compra y uso de los productos.
- Cómo buscan información del sector textil.
- Tiempos de entrega y de inventario.

Algunas estadísticas para la construcción de los perfiles de los segmentos se presentan a continuación:

Tabla 14: Participación de los segmentos

Segmento	N. de Clientes	% Part.	Ventas (Millones)	% Part.	Devoluciones (Millones)	% Part.	U. Operativa (Millones)	% Part.
Industrial MP	280	8,6%	\$ 67.113	40,4%	\$ 4.600	66,3%	\$ 3.783	75,5%
C. Femenina - MP	262	8,0%	\$ 19.046	11,5%	\$ 322	4,6%	\$ 940	18,8%
C. Masculina - MP	305	9,3%	\$ 20.096	12,1%	\$ 610	8,8%	\$ 877	17,5%
C. Dotaciones - MP	431	13,2%	\$ 13.239	8,0%	\$ 248	3,6%	\$ 236	4,7%
C. Infantil - MP	225	6,9%	\$ 4.546	2,7%	\$ 97	1,4%	\$ 161	3,2%
C. Quirúrgica - MP	110	3,4%	\$ 2.179	1,3%	\$ 16	0,2%	\$ 146	2,9%
Comercio PT	509	15,6%	\$ 3.150	1,9%	\$ 173	2,5%	\$ 112	2,2%
Comercio Mayorista MP	120	3,7%	\$ 15.029	9,0%	\$ 245	3,5%	\$ 85	1,7%
G. Superficies PT	17	0,5%	\$ 2.691	1,6%	\$ 138	2,0%	\$ 43	0,9%
Publicidad MP	94	2,9%	\$ 1.070	0,6%	\$ 14	0,2%	\$ -25	-0,5%
C. Ropa deportiva - MP	114	3,5%	\$ 3.309	2,0%	\$ 97	1,4%	\$ -153	-3,1%
Comercio Minorista MP	647	19,8%	\$ 10.657	6,4%	\$ 212	3,1%	\$ -330	-6,6%
Mayoristas PT	153	4,7%	\$ 4.071	2,4%	\$ 171	2,5%	\$ -863	-17,2%
Total	3267	100%	\$ 166.196	100%	\$ 6.943	100%	\$ 5.012	100%

Fuente. COMERTEX S.A.

Un ejemplo de la caracterización de los segmentos se hace mediante la articulación de tecnología y el uso de herramientas estadísticas que facilitan el manejo de la información y se resume en infografías que permiten una mejor visualización de las características del segmento. (Ver figura 33)

Figura 33. Infografía caracterización de un grupo de clientes



Fuente. Autor

El proceso se repite para los 12 segmentos restantes. En el Anexo H: Caracterización de los segmentos se encuentran las estadísticas para cada segmento así como un resumen de la caracterización.

5.2 PROMESA DE SERVICIO

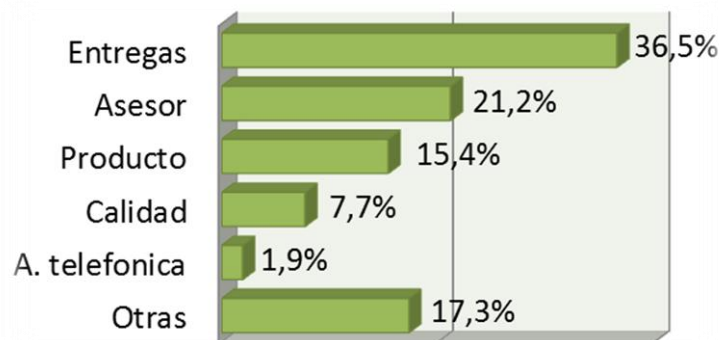
Soportados en la caracterización de cada segmento, por medio de la base de datos transaccional y la encuesta del nivel de servicio, se logró identificar algunos aspectos de mejora para cada criterio de evaluación del nivel de servicio. A continuación se relacionan estos criterios y cada uno de los aspectos de mejora como parte de la construcción de un nivel mínimo o estándar de servicio que COMERTEX S.A. debe ofrecer a sus clientes.

Los aspectos más relevantes relacionados con el servicio a los clientes se presentan a continuación:

- *Asesor comercial*
- *Producto*
- *Entregas*
- *Atención telefónica*
- *Análisis del proceso de venta cruzada para cada segmento*

El análisis parte de la identificación aproximada del 17% de insatisfacción de los clientes y su relación con las causas de esa insatisfacción como se muestra en la figura 34.

Figura 34: Relación causas de insatisfacción



Fuente. Autor

Del 17% de insatisfacción el 36,5% corresponde a la insatisfacción ligada al incumplimiento en las entregas de los pedidos.

A partir del análisis de la encuesta de servicio se da una serie de recomendaciones para cada uno de los ítems presentados anteriormente y están relacionados con los resultados de los criterios de evaluación para cada aspecto. Los criterios evaluados a cada aspecto se encuentran en el Anexo 2: Plantillas de trabajo y encuesta de nivel de servicio.

5.2.1 Asesor comercial.

Los criterios evaluados al asesor comercial son:

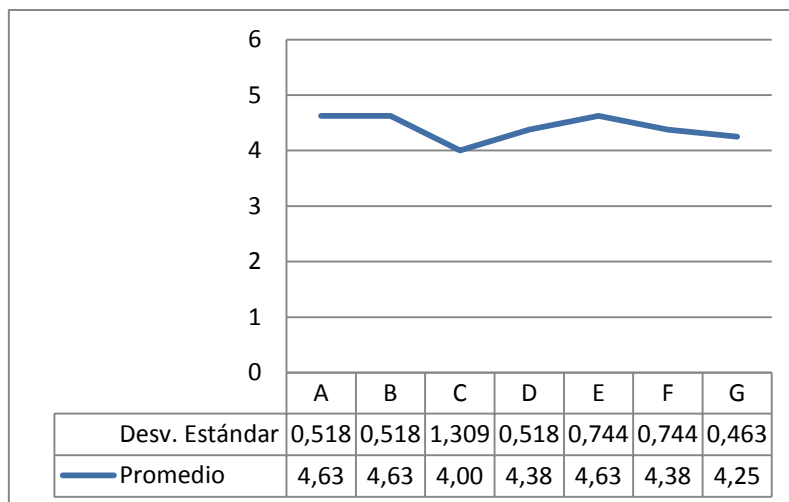
Tabla 15: Criterios evaluados al Asesor Comercial

A. Amabilidad en el servicio ofrecido por el asesor comercial de Comertex S.A.
B. Disposición y disponibilidad del asesor comercial de Comertex S.A. para atender sus requerimientos
C. Velocidad con la que el asesor comercial de Comertex S.A. le responde a sus inquietudes y/o solicitudes
D. Presentación personal del asesor comercial de Comertex S.A.
E. Expresión verbal usado por el asesor comercial
F. Información suministrada por el asesor comercial de Comertex S.A.
G. Conocimiento y dominio del portafolio por parte del asesor comercial de Comertex S.A.

Fuente. COMERTEX S.A.

Los resultados de la evaluación de estos criterios se presentan en la figura 35.

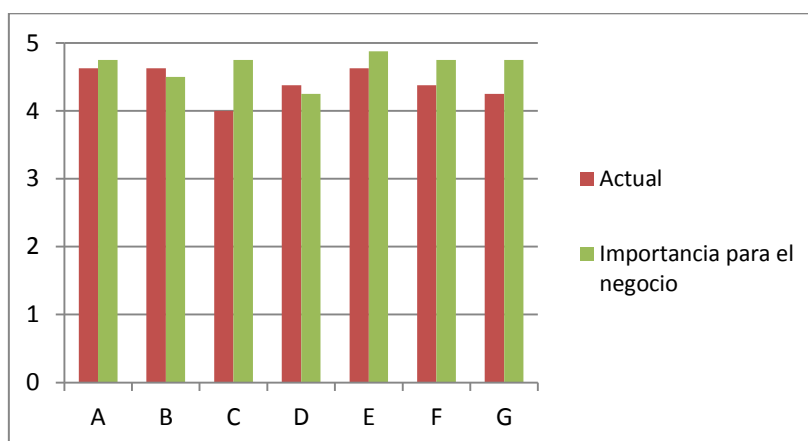
Figura 35: Evaluación promedio a los criterios asociados al Asesor Comercial



Fuente. Autor

En la gráfica anterior se observan algunas posibilidades de mejora en cuanto a los criterios evaluados lo cual se corrobora con la figura 36 en la cual se compara el nivel percibido por el cliente con la importancia que tiene para éste en el desarrollo de la relación comercial.

Figura 36: Comparación del aspecto evaluado y la importancia que tiene para el negocio



Fuente. Autor

Dado este análisis es necesario trabajar en los siguientes criterios:

- Garantizar que el asesor tenga un canal permanente de comunicación con el cliente. El canal de comunicación se puede dar por medio de la página web de la empresa, vía telefónica, visitas o e-mail. El asesor debe registrar el momento exacto en que el cliente realiza el pedido así como el medio por el cual el cliente se comunicó con él.
- El asesor debe guardar información del cliente respecto al estado del cupo de crédito, para determinar si el pedido puede ser facturado.
- Para que las estrategias tengan éxito es necesario tener un plan de capacitación para los asesores respecto a la atención de clientes y un conjunto de beneficios para ellos que deberán ser discutidos con la gerencia.

Un modelo de formulario para el registro de pedidos se presenta en la tabla 13.

Tabla 16. Modelo para el registro de pedidos

Consecutivo _____

Nit: _____ Teléfono _____

Nombre: _____

Fecha: Día: _____ Mes: _____ Año: _____

Medio Comunicación Visita: _____ Teléfono: _____ E-mail: _____ Web: _____

Crédito: Si: _____ No: _____

Cantidad	Nombre producto	Valor unitario	Valor total

Observaciones: campo para registrar la autorización para el envío de productos sustitutos

Fuente. Autor

El análisis se repite para cada uno de los aspectos que siguen y en adelante se detallan las acciones para garantizar un nivel de servicio mínimo o estándar.

5.2.2 Producto.

- COMERTEX S.A. se compromete a garantizar la calidad de cada uno de los productos que envía a sus clientes y a su vez, enviar exactamente lo que el cliente necesita, o previo acuerdo con el cliente enviar productos sustitutos. Se debe registrar el tipo de acuerdo que se realiza con el cliente.
- Así mismo, suministrará el catálogo de productos vía e-mail y en su página web con lo cual se mostrará una posición de compromiso

ambiental. Un departamento de calidad se hará cargo de hacer seguimiento de las propuestas descritas.

- COMERTEX S.A. se compromete a crear un link en la página web donde se muestre la trazabilidad de los productos desde el momento de la compra hasta el momento que se envía el pedido, creara un espacio visible para que el cliente conozca parte del proceso.

5.2.3 Entregas.

- COMERTEX S.A. se compromete a entregar cada uno de los pedidos en un promedio de 3 o menos días, dependiendo del segmento, tiempo que incluye desde que el cliente hace el pedido al asesor hasta que el cliente recibe el pedido y enviar los productos que el cliente solicita en las cantidades que el cliente pide. En caso de incumplimiento COMERTEX S.A. se compromete a generar algún tipo de acción (descuentos especiales).
- En la página web se debe mostrar de manera explícita la política de servicio que tiene COMERTEX S.A. con sus clientes.

5.2.4 Atención quejas y reclamos.

- COMERTEX S.A. ofrecerá atención a quejas y reclamos mediante una línea de atención telefónica gratuita y tendrá un espacio claramente definido para este fin en su página web. Se establecerá un tiempo mínimo de respuesta y un tiempo máximo dependiendo de la criticidad de la queja o reclamo. En este aspecto es muy importante que se construya un espacio para que el cliente registre sus apreciaciones acerca de las devoluciones que se generan. En el caso de la atención telefónica COMERTEX S.A. debería tener la posibilidad de registrar el número telefónico del cliente y si no se pudo atender el cliente por congestión en las llamadas en el mínimo tiempo posible regresar la

llamada y tomar nota de la sugerencia o queja. Las quejas y reclamos serán atendidas por el departamento de calidad.

5.2.5 Información.

- COMERTEX S.A. se compromete a suministrar información del sector textil, ferias o eventos de la compañía a sus clientes con una periodicidad de 15 días, extenderá invitaciones especiales o gratuitas a los clientes VIP o a aquellos que a la fecha superen su promedio histórico de ventas. Para mayor detalle ver Anexo I: Promesa de servicio.

Cada aspecto de posible mejora fue identificado gracias a los datos arrojados por la encuesta y sobre la cual, en primera instancia, se analizaron las deficiencias en el servicio indistintamente del segmento al cual pertenecen los clientes.

5.3 Estrategias de servicio por segmento

A continuación se detallan algunos aspectos que permiten diferenciar el servicio para cada segmento y otros aspectos en los cuales COMERTEX S.A. debe garantizar la mejora en el servicio. Existe información que es igual para algunos segmentos, caso(s) en el(los) cual(es) la(s) causa(s) que los generan son diferentes. Se recomienda la verificación de cada una de las hipótesis presentadas. Las estrategias se plantean mediante la descripción de las características más relevantes para cada segmento.

En la tabla 17 se muestra el análisis comparativo de los segmentos a partir de variables de estudio que permiten establecer algunas características y comportamientos similares y en otros reflejar las diferencias para facilitar la generación de estrategias diferenciadas de servicio.

Tabla 17: Comparación de segmentos

Criterios de evaluación	C. May MP	C. Min MP	Confecciones					
			Dotaciones	Femenina	Infantil	Masculina	Quirúrgica	Ropa deportiva
N. de clientes	120	647	431	262	225	305	110	114
Ventas (%participación)	9,1%	6,4%	8,0%	11,5%	2,7%	12,1%	1,3%	2,0%
Líneas producto (representativas)	Tejido Plano (82%)	No tejidos (95%) Tejido Plano Tejido Punto	Tejido Plano(82%)	Tejido Plano (94%)	Tejido Plano Tejido Punto	Tejido Plano(94%)	No Tejidos	Tejido Punto(90%)
Grupo de producto (representativos)	Indigos	Gran número de grupos	Driles Indigo	Indigos	Gran número de grupos	Indigos	Bonlam	Elastomeros Deportivos
Devoluciones (% participación)	3,5%	3,1%	3,6%	4,6%	1,4%	8,8%	0,2%	1,4%
Cartera (Días para el pago)	89	71	82	110	109	117	48	62
Pedidos	Cada 10 días Valor: 3,5 millones	Cada 32 días Valor: 1,5 millones	Cada 31 días Valor: 2,6 millones	Cada 63 días Valor: 1,8 millones	Cada 37 días Valor: 2 millones	Cada 53 días Valor: 9,7 millones	Cada 37 días Valor: 2 millones	Cada 28 días Valor: 2,2 millones
Utilidad Operacional (% participación)	1,7%	-6,6%	4,7%	18,7%	3,2%	17,5%	2,9%	-3,0%
Número de encuestados	8	61	41	23	12	37	4	3
Satisfacción	62,5%	85,2%	90,2%	82,6%	81,8%	89,5%	100,0%	100,0%
Aspectos importantes a la hora de la compra	Precio Calidad del producto	Calidad del producto Precio	Calidad del producto Precio	Calidad del producto Precio	Precio Calidad del producto	Calidad del producto Precio	Calidad del producto Precio	Calidad del producto Precio Tiempo de entrega
Razones de compra	Producción Distribución	Comercialización Confección	Confecciones	Confección Precio	Confección	Confección Comercialización	Confección	Confección
Razones de uso	Confección prendas de vestir, venta	Ropa dama, almohadas, Toallas, Zapatos	Uniformes, Dotaciones, Ropa deportiva	Jeans, terminación de prendas, Comercialización	Ropa para bebe, pijamas	Pantalones, camisas	Ropa hospitalaria	Ropa deportiva y uniformes
Busqueda de información	R. de ventas Internet	R. de ventas Internet Distribuidores	Internet Distribuidores Terceros	Distribuidores Internet	R. de ventas Internet	Distribuidores R. de ventas	Distribuidores R. de ventas	Internet Distribuidores
Entregas	Actual= 5 días Ideal= 2 días	Actual= 4 días Ideal= 2 días	Actual= 3 días Ideal= 2 días	Actual= 5 días Ideal= 3 días	Actual= 2 días Ideal= 3 días	Actual= 4 días Ideal= 3 días	Actual= 5 días Ideal= 4 días	Actual= 1 días Ideal= 1 días
Inventario	No les gusta	5 días	2 días	8 días	2 días	2 días	45 días	No les gusta

Fuente. Autor

Tabla 18: Comparación de segmentos

Criterios de evaluación	Industrial	Publicidad	Producto Terminado		
			Grandes Superficies	Mayoristas	Comercio
N. de clientes	280	94	17	153	509
Ventas (%participación)	40,4%	0,6%	1,6%	2,4%	1,9%
Líneas producto (representativas)	Algodón No Tejidos Tejido Plano	No Tejidos (95%) Tejido Punto Tejido Plano	Hogar (97,5) Calzado	Calzado (95,7) Hogar	Hogar (80%) Calzado
Grupo de producto (representativos)	Bonlam Importadas	Bonlam Punto sencillo	Baño Infantil Descanso	Infantil Calzado deportivo Deportivo	Baño Alcoba Infantil
Devoluciones (% participación)	66,2%	0,2%	2,0%	2,5%	2,5%
Cartera (Días para el pago)	62	43	114	88	78
Pedidos	Cada 57 días Valor: 2,7 millones	Cada 63 días Valor: 1,9 millones	Cada 2 días Valor: 547 millones	Cada 29 días Valor: 2,2 millones	Cada 103 días Valor: 1,8 millones
Utilidad Operacional (% participación)	75,5%	-0,5%	0,9%	-17,2%	2,2%
Número de encuestados	28	3	3	8	74
Satisfacción	92,9%	100,0%	100,0%	87,5%	68,9%
Aspectos importantes a la hora de la compra	Calidad del producto Precio Tiempo de entrega	Calidad del producto Precio	Calidad del producto Diseño del producto Tiempo de entrega Asesoría comercial Servicio pos venta	Calidad del producto Precio	Calidad del producto Diseño del producto Precio
Razones de compra	Confección de gran de volumen	Producción, venta	Calidad Precio	Precio Rapidez para hacer negocios	El cliente lo pide Precio
Razones de uso	Almohadas, lencería, cubrelechos, cojines	Camisetas, ropa desechable y publicidad	Comercialización	Comercialización	Comercialización
Busqueda de información	Terceros Internet	Internet Distribuidores	R. de ventas Internet	R. de ventas Internet Revistas especializadas Eventos de la compañía	R. de ventas Internet
Entregas	Actual= 4 días Ideal= 3 días	Actual= 2 días Ideal= 1 días	Actual= 9 días Ideal= 5 días	Actual= 4 días Ideal= 3 días	Actual= 7 días Ideal= 3 días
Inventario	13 días	1 días	2 días	4 días	7 días

Fuente. Autor

❖ Comercio Mayorista MP

- Ofrecer velocidad de respuesta por parte del asesor: entre 2 y 5 horas para atender y solucionar los requerimientos del cliente.
- Mejor precio y variedad en el producto: se debe analizar el proceso de compra y definir la “canasta” de productos.
- Tiempo de entrega: 2 días.
- Mejorar la atención telefónica en cuanto al tiempo de espera, tipo de solución y al servicio de la atención telefónica: la solución deberá estar entre 8 y 12 horas luego de ser registrada la queja y se le dará al cliente un número de radicado.
- Impulsar la asistencia a ferias y eventos del sector textil: suministrar información cada 15 días.

❖ Comercio Minorista MP

- Aumentar el número de visitas al cliente por parte del asesor: 1 visita cada 15 días.
- Mejorar el precio, variedad y disponibilidad del producto: analizar el proceso de compra y definir la canasta de productos.
- Tiempo de entrega: 2 días.
- Ofrecer servicio post-venta.
- Atención telefónica debe ser rápida, dar soluciones e informar correctamente: información inmediata o de acuerdo a la criticidad.
- Impulsar las ferias y eventos, los clientes dicen que éstos los motivan a comprar: suministrar información cada 15 días y orientarlos en las tendencias del sector textil.
- Para este segmento es más importante la calidad del producto que el precio.
- Los representantes de ventas deben informar correctamente a los clientes sobre el sector textil ya que es el medio más utilizado seguido de internet.

- Aumentar el valor promedio de cada pedido y disminuir la frecuencia entre pedidos, es decir, que pidan más en valor y menos frecuentemente: descuentos por volumen de compra o que el cliente cancele el flete.

❖ **Confecciones dotaciones MP**

- Aumentar el número de visitas al cliente.
- Mejorar precio y disponibilidad del producto: ofrecer descuentos por pronto pago y realizar el estudio de demanda que permita garantizar la disponibilidad de los productos para este segmento.
- Tiempo de entrega: 1 día.
- Atención telefónica: atención inmediata y establecer tiempo de solución de acuerdo a la criticidad del asunto.
- Segmento que es más sensible a la calidad del producto que al precio.
- Busca información a través de la web: mantener la web de COMERTEX S.A. actualizada con información personalizada para este segmento.

❖ **Confecciones femenina MP**

- El asesor debe responder rápidamente a las inquietudes, brindar información y tener dominio del portafolio de producto así como aumentar el número de visitas al cliente.
- Mejorar la percepción que el cliente tiene sobre el precio, la disponibilidad, la variedad y el diseño del producto.
- Tiempo de entrega: 2 días y servicio post-venta.
- Prestar atención a quejas y reclamos, disminuir el tiempo de espera y mejorar la calidad de la información que se suministra.
- Ofrecer calidad y precio en cada producto.
- Retroalimentación con distribuidores ya que el segmento se informa con ellos en primera instancia y luego con el uso de internet: crear alianzas

con distribuidores para la retroalimentación con información del sector y de los clientes así como ofrecerla en la página web de COMERTEX S.A.

❖ **Confecciones infantil MP**

- El asesor comercial debe ser rápido en responder a las solicitudes del cliente y ofrecer excelente información.
- El producto debe tener buen precio, variedad, diseño y disponibilidad: análisis del proceso de compra y de la canasta de productos.
- Tiempo de entrega: 2 días.
- Mejorar en tiempos de espera en la atención telefónica.
- Impulsar el desarrollo de ferias y eventos ya que para este segmento es una motivación para comprar.
- Lo más importante a la hora de la compra es el precio seguido de la calidad del producto.
- Ofrecer información del sector textil a través de internet y representantes de ventas.

❖ **Confecciones masculina MP**

- El asesor debe ser amable, mostrar disposición y disponibilidad, ser rápido en la respuesta al cliente, brindar información de calidad, conocer con detalle el portafolio de producto y visitar frecuentemente al cliente.
- El producto debe evidenciar buena relación entre precio, variedad y disponibilidad.
- Tiempo de entrega: 2 días.
- Ofrecer servicio post-venta.
- La atención telefónica debe orientar las quejas y los reclamos así como disminuir el tiempo de espera en la línea.
- Las ferias motivan la compra de quienes asisten a ellas.
- Los clientes son más sensibles a la calidad que al precio.

- Mantener buena relación con los distribuidores pues son ellos quienes frecuentemente informan a los clientes sobre el sector textil.

❖ **Confecciones quirúrgica MP**

- Es importante mantener el nivel de servicio del segmento en cada aspecto como asesor comercial, producto, entrega, atención telefónica y adelantarse a posibles necesidades que faciliten la generación de valor.
- El asesor es fundamental en el servicio.
- Son sensibles por igual al precio y la calidad del producto.
- Tiempo de entrega: 1 día (24 horas).
- Debe suministrarse información vía internet.
- Asesoría y facilidades para el manejo de inventarios.
- Negociar tiempo de entrega.

❖ **Confecciones ropa deportiva MP**

- Mantener el nivel de servicio, adicionar disponibilidad del producto.
- Aumentar el conocimiento del segmento.
- Tiempo de entrega: 1 día (24 horas).
- Sensibles al precio, calidad y tiempos de entrega.
- Ofrecer facilidades de pago.

❖ **Industrial MP**

- El asesor deber suministrar buena información al cliente y aumentar el número de visitas.
- El producto debe ofrecer buena relación entre precio, variedad y disponibilidad.
- Mejorar las entregas en cuanto a fecha de entrega y servicio post-venta.
- Se debe reducir el tiempo de espera en la atención telefónica así como optimizar la información que se da al cliente.

- Tiempo de entrega: 2 días.
- Incentivar la asistencia a ferias y eventos ya que estos los incentivan a comprar.
- Clientes del segmento más sensibles a la calidad que al precio.
- Suministrar información del sector textil preferiblemente a través de internet.
- Asesoría para manejo de inventarios.

❖ **Publicidad MP**

- Aumentar el número de visitas por parte del asesor comercial.
- Asegurar que el cliente perciba la relación precio, variedad, diseño y disponibilidad para el producto.
- Mejorar el cumplimiento en la entrega y cantidades solicitadas por el cliente.
- Tiempo de entrega: 1 día (24 horas).
- La atención telefónica debe asegurarse de ofrecer soluciones que satisfagan los requerimientos del cliente.
- Ofrecer calidad del producto por encima del precio.
- Información suministrada principalmente a través de internet.

❖ **Grandes superficies**

- El asesor debe aumentar el número de visitas, mostrar disposición y disponibilidad al momento del contacto con el cliente y ser rápido en la atención de sus necesidades.
- El producto debe ofrecer variedad y disponibilidad.
- Las entregas deben ser cumplidas en cuanto a fecha de entrega y cantidades entregadas con respecto a la facturación, el producto debe llegar en perfectas condiciones y ofrecer servicio post-venta.
- Tiempo de entrega: No aplica.

- Se deben atender de manera oportuna las quejas y reclamos y reducir el tiempo de espera a dichas solicitudes.
- Sensibilidad a la calidad del producto seguido de diseño, tiempos de entrega y servicio post-venta; asumen buen precio por su poder de negociación.

❖ **Mayoristas PT**

- El asesor debe asegurar que la información suministrada al cliente satisface sus requerimientos.
- Ofrecer variedad y disponibilidad de producto.
- Mejorar el cumplimiento en la entrega de pedidos y adicionar buen servicio post-venta.
- Tiempo de entrega: 1 días.
- Mayor sensibilidad a la calidad que al precio del producto.
- El representante de ventas debe estar bien informado acerca del sector textil ya que los clientes buscan información a través de él seguido de internet y revistas especializadas.

❖ **Comercio PT**

- El asesor debe mostrar disposición y disponibilidad, rapidez para informar, tener total dominio del portafolio y visitar más frecuentemente al cliente.
- Relación precio, calidad y variedad para el producto.
- Mejorar tiempos de entrega y asegurar que las cantidades corresponden a lo solicitado o facturado.
- Tiempo de entrega: 2 días.
- Reducir tiempo de espera en la atención telefónica y recibir de forma adecuada las quejas y reclamos.
- Motivar la asistencia a ferias o eventos ya que incentiva la compra.
- Sensibilidad a la calidad del producto seguido del diseño.

- El representante de ventas debe informarle sobre el sector textil.

5.4 CLIENTES VIP

El proceso de caracterización de clientes permitió la estandarización y formalización del concepto de cliente **VIP** en el cual COMERTEX S.A. definió como criterio “*cliente VIP es aquel que genere utilidad operativa al final del año por valor igual o superior a \$12 millones, sin tener en cuenta aquellos clientes que compran algodón*”.

En este sentido y dado el análisis de la tabla 19 en la cual se observa la fuerte participación de los clientes VIP en cuanto a ventas y utilidad operacional se hace necesario la identificación y el trabajo personalizado con este tipo de clientes siendo éstos quienes generan mayor utilidad operativa y mayor rentabilidad. En el caso del tipo de clientes *Algodón* solo compran este insumo, su rentabilidad es menor y hace parte de un negocio marginal de la empresa, por tanto, solo se tendrá en cuenta para el análisis comparativo completo.

Tabla 19: Tipos de clientes

Tipo de Clientes	Ventas	% Part.	Utilidad Operacional	% Part.	N. de Clientes	% Part.
VIP	\$ 46.223.117.706	27,8%	\$ 4.286.124.824	50,4%	109	3,3%
NO VIP	\$ 57.583.829.381	34,7%	\$ 673.967.017	7,9%	3139	96,3%
Algodón	\$ 62.376.244.453	37,5%	\$ 3.544.157.643	41,7%	10	0,3%
TOTAL	\$ 166.183.191.540		\$ 8.504.249.484		3258	

Fuente. Autor

La identificación de los VIP también se realizó en cada uno de los segmentos buscando conocer más sobre ellos y su participación en el negocio. (Ver tabla 20)

Al respecto, se dan algunos lineamientos que permitirían tener un concepto más robusto y algunos servicios complementarios que permitan la

diferenciación del servicio y definir una estrategia clara para potencializar clientes.

Tabla 20: VIP por segmentos

Segmento	N. de Clientes	Ventas	% Part.	Utilidad Operativa	% Part.	Cartera (Días)
C.Femenina	25	\$ 12.420.679.109	26,9%	\$ 1.503.240.330	35,1%	158
C. Masculina	38	\$ 12.333.482.239	26,7%	\$ 1.238.236.288	28,9%	146
Mayorista MP	10	\$ 11.392.091.372	24,6%	\$ 746.553.587	17,4%	118
C. Dotaciones	15	\$ 5.131.765.259	11,1%	\$ 369.551.069	8,6%	139
C. Ropa deportiva	4	\$ 1.465.120.763	3,2%	\$ 81.698.777	1,9%	107
Minorista MP	7	\$ 1.356.069.943	2,9%	\$ 133.128.599	3,1%	89
G. Superficies PT	2	\$ 760.222.921	1,6%	\$ 55.735.236	1,3%	58
C. Quirúrgica	3	\$ 462.847.483	1,0%	\$ 61.127.015	1,4%	130
C. Infantil	2	\$ 436.487.897	0,9%	\$ 39.082.660	0,9%	132
Industrial MP	2	\$ 332.359.280	0,7%	\$ 35.672.690	0,8%	108
Publicidad MP	1	\$ 131.991.440	0,3%	\$ 22.098.573	0,5%	70
TOTAL	109	\$ 46.223.117.706		\$ 4.286.124.824		

Fuente. Autor

Por ejemplo, el concepto de cliente VIP debería incluir aspectos como:

- Rentabilidad
- Exclusividad
- Ventas
- Cantidad de pedidos

En cuanto a las estrategias que se podrían ofrecer a este tipo de clientes están:

- Conocimiento de que son clientes VIP
- Atención personalizada (Directamente por los gerentes de área)
- Ofrecer servicios gratuitos (envío de mercancía)
- Invitación a eventos especiales

5.5 BASE DE DATOS

En el desarrollo del proceso se encontró una fuerte limitación en cuanto al tipo de datos que se almacenan y la pureza éstos. Básicamente los datos corresponden a datos transaccionales estructurados de forma particular *CLIENTE* a *PRODUCTO*, es decir, los datos asociados a la factura de un cliente en particular se encuentran relacionados por referencia de producto como un registro, por ejemplo, para una factura que incluya 2 referencias de producto se tienen 2 registros independientes. En el caso de la plantilla 10, que muestra ventas por factura, no se tiene información de los productos enviados.

Las plantillas suministradas son pre-establecidas y no fueron creadas de acuerdo a las necesidades de este proyecto. A continuación se muestra una aproximación a los datos que se deberían capturar para hacer más efectiva la gestión de las relaciones con el cliente a través de un conocimiento profundo de éstos.

- Registrar exactamente lo que el cliente solicita en cada pedido.
- Enlaces de interés para el cliente cuando visita la página web de COMERTEX S.A.
- Tiempos de entrega, desde que el cliente realiza el pedido hasta que lo recibe.
- Asistencia a ferias, showroom, cafés de negocio y qué tipos de producto compran.
- Pago de facturas dentro del tiempo de crédito que fue otorgado.

Entre otros, estos son algunos datos para conocer a los clientes, por tanto, es fundamental que se pueda capturar cualquier contacto que se tenga con ellos.

Otro aspecto muy importante es poder garantizar la pureza de la información para la cual se deben crear protocolos para ingreso de datos al sistema y repositorios de información. En este sentido se puede automatizar el proceso

para que cada área, previa identificación de sus necesidades, reciba los datos e información y evitar que tengan contacto directamente con la base de datos.

CONCLUSIONES

- COMERTEX S.A. no cuenta con una base de datos sólida y diversificada, aspecto que limita el uso de técnicas de minería de datos para la generación de nuevo conocimiento en beneficio de generar valor en cada uno de sus servicios.
- La información y los datos deberían ser los mismos para cada área de la empresa y estar ligados al ERP. Se presenta dificultad con los datos por la forma en que se ingresan los registros al sistema.
- Las devoluciones representan cerca del 4% de las ventas totales y están concentradas, principalmente (85%), en tres criterios: error en la toma o digitación del pedido, error en el despacho y producto de calidad no conforme.
- El 80% de los clientes de COMERTEX S.A. se encuentran ubicados en 7 de los 13 segmentos pero en cuanto a la participación de ventas solo el segmento Industrial MP alcanza el 40% de las ventas totales, de ese 40% el 37,5% de las ventas corresponde a la venta de algodón y cuya cantidad de clientes es el 3% del total de clientes de la empresa.
- Cuatro (4) de los 13 segmentos registraron déficit en la utilidad operativa del año 2011, corresponde a 886 clientes, por cerca del 27% de la utilidad operativa total; en este aspecto el segmento Industrial MP representa más del 75% de la utilidad operativa lo cual muestra una fuerte dependencia, específicamente de los clientes que compran algodón.

- Los clientes VIP representan el 50,4% de la utilidad operacional y en cuanto a cantidad de clientes representan el 3,3% lo cual evidencia la necesidad de construir estrategias especiales de servicio.
- El conocimiento del cliente juega un papel fundamental en la generación de nuevos negocios y en este aspecto el uso de tecnologías de la información y herramientas estadísticas avanzadas permiten la obtención de información que de otra manera sería muy difícil de encontrar.
- El estudio del nivel de servicio mostró que COMERTEX S.A. se encuentra por encima del estándar de servicio, sin embargo existen oportunidades de mejora específicas en cada uno de los aspectos evaluados. Gran parte de la insatisfacción de los clientes está relacionada con el incumplimiento y la demora en la entrega de los pedidos.
- El uso de redes bayesianas permitió la identificación de relaciones entre las variables del servicio así como relaciones de independencia/dependencia entre las líneas de producto que se venden en COMERTEX S.A.; a partir de esto se construyó la promesa de servicio.
- La construcción del perfil de los clientes de acuerdo al segmento en el que se encuentran permitió la identificación de oportunidades de negocio y de mejora. Facilitó la generación de una promesa de servicio y la necesidad de continuar con el estudio profundo de las necesidades del cliente.

RECOMENDACIONES

- COMERTEX S.A. debe elaborar un protocolo para el ingreso de datos al ERP y para el tratamiento de éstos que evite el deterioro y mantenga la pureza de la información. Junto con esto se debe crear un repositorio de datos al cual se pueda recurrir para la búsqueda de información.
- La información debe ser la misma para cada área de la empresa, en este sentido se deben realizar auditorías de control de datos y manejo de la información. Definir qué tipo de información requiere cada área de la empresa y que sea el sistema quien se la suministre diariamente evitando la manipulación de los datos.
- Un enfoque orientado al cliente es fundamental para la generación de valor, por tanto, COMERTEX S.A. debe capturar la mayor cantidad de datos acerca de los movimientos y gustos de los clientes. Por ejemplo, se podrían capturar datos acerca de los productos que al cliente le interesan cuando visita la página web o los datos de aquellos clientes que asisten a las ferias, eventos o aquellos que responden de manera positiva a las ofertas y promociones.
- Realizar un análisis del proceso de compra y junto a este un análisis de la canasta de compra por segmento lo cual permitiría incluir nuevas estrategias con productos de interés o productos sustitutos que garanticen la existencia de productos de interés para el cliente.
- Optimizar el uso de las tecnologías de la información para generar cercanía con el cliente, atender de manera oportuna sus necesidades y realizar marketing en línea. Esto se logra en primera instancia con el adecuado manejo de la página web a través de catálogos actualizados, información de interés para el cliente, enlaces específicos para la

gestión de reclamaciones, para conocer su estado financiero o simplemente para compartir información entre clientes.

- A partir de los resultados obtenidos definir una estrategia mínima de servicio que incluya los aspectos que se relacionan en el capítulo 5. Realizar una nueva encuesta de servicio incluyendo las preguntas que se sugieren en el mismo capítulo. Con los resultados de esta nueva encuesta re-definir la estrategia diferenciada por segmento para la cual se da un punto de partida soportados en los datos de la encuesta de servicio del 2011.
- Como parte de la generación de valor, es muy importante que COMERTEX S.A. construya diferentes procesos como gestión de devoluciones, gestión de reclamaciones, análisis de venta cruzada, correo e información dirigida a clientes específicos y predicciones de ventas.
- Validar cada una de las hipótesis planteadas en este trabajo de grado mediante la aplicación de nuevas encuestas de servicio o la captura, almacenamiento y tratamiento estadístico de los nuevos datos que se siguieren en el presente trabajo de grado.
- Basados en las dos recomendaciones anteriores, las tendencias en mejores prácticas de servicio y marketing están orientadas hacia el uso de tecnologías de información que permitan optimizar estos aspectos y generar información actualizada y en tiempo real, por tanto, la gestión de marketing y servicio de COMERTEX S.A. deberían estar integradas en sistemas en línea como por ejemplo una red social (tipo Facebook, LinkedIn) en la cual se tenga acceso a información directa e inmediatamente para conocer las necesidades de los clientes.

BIBLIOGRAFÍA

- ÁNGELES, María; SANTILLÁN, Angélica. Minería de datos: Concepto, características, estructura y aplicaciones. [en línea]. [consultado 5 de abr. 2012]. Disponible en <<http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>>
- BEN-GAL, Irad. Bayesian networks. Encyclopedia of statistics in Quality & Reliability, Wiley & Sons (2007)
- BRITOS, Paola. Objetivos de negocio y procesos de minería de datos basados en sistemas inteligentes. [en línea]. [consultado 5 abr. 2012]. Disponible en <<http://laboratorios.fi.uba.ar/lsi/R-rtis-7-1-2005-26-29.pdf>>
- CASTAÑEDA, José; RODRIGUEZ, Miguel. La minería de datos como herramienta de marketing: Delimitación y medidas evaluación del resultado. [en línea]. [consultado 5 abr. 2012]. Disponible en <http://www.ugr.es/~jalberto/Investigacion/Casta_ed4.pdf>
- COBOS, Carlos et al. CMIN- herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos. Revista ingeniería e investigación. Vol. 30, No 3 (2010). [consultado 5 abr. 2012]
- CRAVERO, Ania; SEPÚLVEDA, Samuel. Aplicación de la minería de datos para la detección de anomalías: un caso de estudio. WorkShop EIG 2009. [en línea]. [consultado 30 abr. 2012]. Disponible en <http://ceur-ws.org/Vol-558/Art_8.pdf>
- DIAZ, José Luis. Utilización de técnicas avanzadas en el tratamiento y manejo de datos. Aplicación a la gestión de sistemas de abastecimiento de agua. Valencia, España. 2010, 382 p. Universidad politécnica de Valencia. Departamento de ingeniería hidráulica y medio ambiente.

- E.W.T. Ngai et al. Application of data mining techniques in customer relationship management: A literature review and classification. [en línea]. [consultado 4 may. 2012]. Disponible en <<http://www.elsevier.com/locate/eswa>>
- GARCÍA TORRES, Miguel. Aplicación de técnicas metaheurísticas en minería de datos. Tenerife, España 2006/07, 202 p. Universidad de la laguna. Serie tesis doctorales.
- GARRIDO, Lluís; LATORRE, José. Aplicaciones empresariales de Data Mining. QÜESTIIO, vol. 25, 3, 2001, p. 499-508.
- HERNÁNDEZ, José; RAMIREZ, María y FERRI, César. Introducción a la Minería de Datos : Madrid. Pearson Education, 2004. 680 p.
- MARCANO AULAR, Yelitza Josefina y TALAVERA PEREIRA, Rosalba. Minería de Datos como soporte a la toma de decisiones empresariales. Vol. 23, p. 104-118. (2007) [en línea]. [consultado 5 abr. 2012]. Disponible en <http://www.scielo.org.ve/scielo.php?pid=S101215872007000100008&script=sci_arttext>
- MARTINEZ, Francisco. Optimización mediante técnicas de minería de datos del ciclo de recosido de una línea de galvanizado. Logroño, España. 2003, 644p. Tesis doctoral (Doctorado en Ingeniería Industrial). Universidad de la Rioja. Departamento de Ingeniería Mecánica.
- MOLINA, Luis. Data mining: Torturando a los datos hasta que confiesen, 2002.[en línea]. [consultado 5 abr. 2012]. Disponible en <<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>>

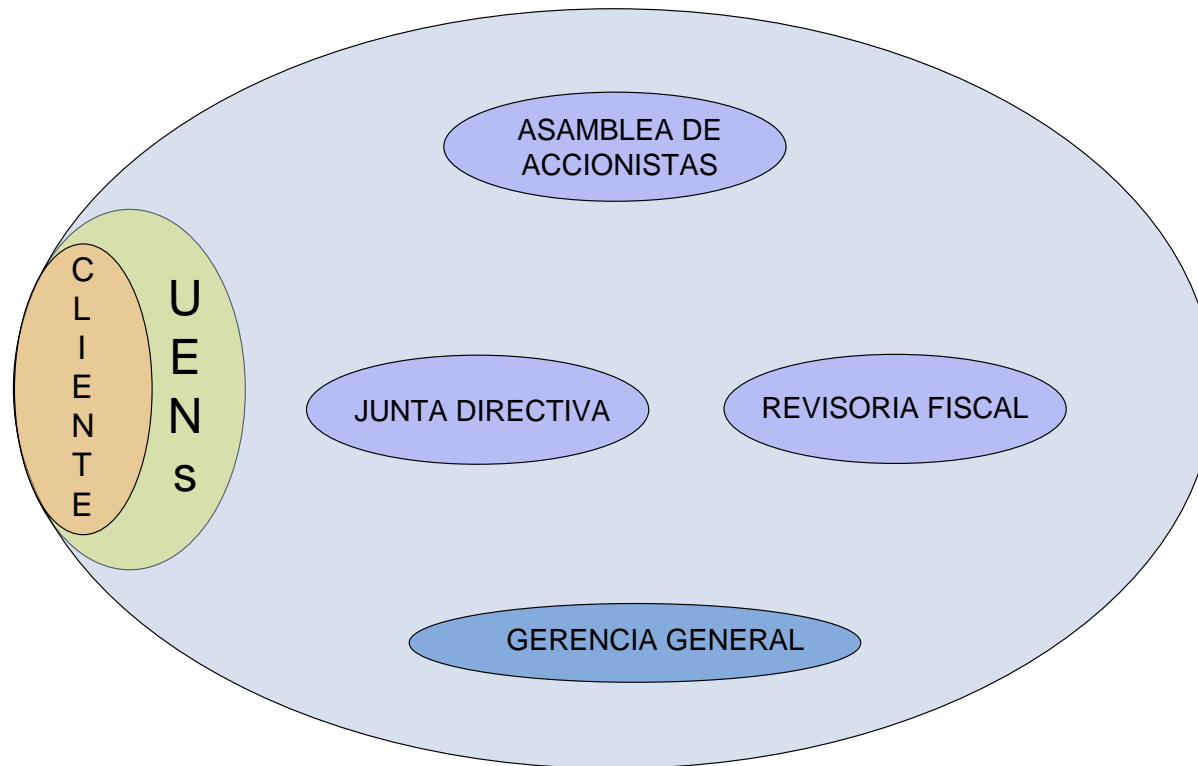
- OLSON, David. Data mining in business services. 2006[en línea].[consultado 5 abr. 2012]. Disponible en <<http://www.springerlink.com/content/x1k175503l007765/>>
- PICÓN, Eduardo; VARELA, Jesús. Segmentando mercados con análisis conjunto. Una aplicación al sector turístico. Psicothema. Vol. 12, suplemento 2 (2000).[en línea]. [consultado 5 abr. 2012]. Disponible en <<http://www.psicothema.com/pdf/603.pdf>>
- PRESSER CARNE, Cynthia. Data mining: Descubriendo información oculta. Argentina. El Cid, 2009.
- REYES, José; GARCÍA, Rodolfo. El proceso de descubrimiento de conocimiento en las bases de datos. En : Ingenierías. Vol. 3, No 26 (2005). [consultado 5 abr. 2012].
- RIQUELME, José; RUIZ, Roberto y GILBERT, Karina. Minería de datos: Conceptos y tendencias. En : Inteligencia artificial, Revista iberoamericana de inteligencia artificial. No 29 (2006). [consultado 5 abr. 2012]
- ROCHE, Ariel. Árboles de decisión y series de tiempo. 2009, 83p. Tesis de maestría en ingeniería matemática. Facultad de ingeniería, UDELAR
- ROCHE, Francisco. Métodos para obtener conocimiento utilizando redes bayesianas y procesos de aprendizaje con algoritmos evolutivos. 2002, 98p. Memoria de para superar la fase de investigación del programa de doctorado. Departamento de lenguajes y sistemas informáticos, Universidad de Sevilla, España.
- ROSADO, Alveiro; RICO, Dewar. Inteligencia de negocios: Estado del arte. En : Red de Revistas Científicas de América Latina, el Caribe, España y Portugal. Vol. 16, No 44 (2010).

- RYGLIELSKI, Chris; WANG, Jyun-Cheng; YEN, David C. Data mining techniques for customer relationship management. [en línea]. [consultado 4 may. 2012]. Disponible en <<http://www.huaat.net/download/DMtechniques.pdf>>
- SHORTLAND, R; SCARFE, R. Data Mining applications in BT. En : BT Technology journal. Vol. 25, Nos 3 y 4 (2007). [consultado 30 abr. 2012]
- TSIPTISIS, Konstantinos; CHORIANOPOULOS, Antonios. Data mining techniques in CRM: Inside customer segmentation. Gran Bretaña. Wiley, 2009. 373 p.
- URIBE, Iván; JIMENEZ, Claudia. Hacia una metodología para la selección de técnicas de depuración de datos. Revista avances en sistemas e información, Vol. 6 No. 1 (2009).
- VÉLEZ, Mario César; MONTOYA, José Alejandro. Metaheurísticos: Una alternativa para la solución de problemas combinatorios en administración de operaciones. Revista EIA, Número 8, p. 99-115. Diciembre 2007.

ANEXOS

Anexo A: ESTRUCTURA ORGANIZACIONAL COMERTEX S.A.

COMERTEX S. A.



Anexo B. PLANTILLAS DE TRABAJO Y ENCUESTA DEL NIVEL DE SERVICIO

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	NIT	NOMBRE	DEPARTAMENTO	CIUDAD	SEGMENTO	Algodón	N. DE SEGM	VIP	Tipo de Cliente	VIP (Si-No)	N. DE VIP	VALOR DE LA	VENDEDOR	DESPACHOS	VEN
2	860002017	CHAIM PEISACH Y CIA. HILAND	DISTRITO CAPITAL	BOGOTA D.C	Industrial M	Si	9	10	Algodoneros	No	120	1089346611	RICARDO VIV	27	153
3	800225401	HILOS DE MOSQUERA S.A. .	CUNDINAMARCA	MOSQUERA	Industrial M	Si	9	10	Algodoneros	No	120	1343260353	RICARDO VIV	18	69
4	811034061	FASHION KAVARISS LTDA .	RISARALDA	DOS QUEBRA	Confeccione	No	4	1	VIP	Si	1	4410000	PERNIA ZULE	62	68
5	860525814	TEXTILES MIRATEX S.A.S .	DISTRITO CAPITAL	BOGOTA D.C	Industrial M	Si	9	10	Algodoneros	No	120	9378000	RICARDO VIV	7	61
6	811011961	NOVARUTA S.A.S. .	ANTIOQUIA	CALDAS	Industrial M	Si	9	10	Algodoneros	No	120	447957620	RICARDO VIV	31	60
7	800048373	CONSORCIO ABUCHAIBE S.A. .	ATLANTICO	BARRANQUI	Industrial M	Si	9	10	Algodoneros	No	120	0	RICARDO VIV	13	18
8	811041841	C.I. MAXIS TEXTIL S.A. .	NORTE DE SANTAN	CUCUTA	Comercio M	No	1	1	VIP	Si	2	81280033,1	CARLOS GIO	66	36
9	860064081	HILANDERIAS UNIVERSAL S.A.S	DISTRITO CAPITAL	BOGOTA D.C	Industrial M	Si	9	10	Algodoneros	No	120	1471390516	RICARDO VIV	2	23
10	890900193	TEJIDOS DE PUNTO LINDALAN	ANTIOQUIA	MEDELLIN	Industrial M	Si	9	10	Algodoneros	No	120	177774750	RICARDO VIV	31	115
11	860036892	RODRIGUEZ FRANCO Y CIA SCS	DISTRITO CAPITAL	BOGOTA D.C	Grandes Sup	No	11	1	VIP	Si	18	22517007	LUZ MARINA	64	6
12	816007367	CREACIONES J.A EU .	RISARALDA	DOS QUEBRA	Confeccione	No	6	1	VIP	Si	3	0	PERNIA ZULE	22	11
13	60331890	CARMEN ANYUL CASTELLANOS	NORTE DE SANTAN	CUCUTA	Comercio M	No	1	1	VIP	Si	6	62252240	CARLOS GIO	28	10
14	98556431	ALEXANDER HINCAPIE MARIN	ANTIOQUIA	MEDELLIN	Confeccione	No	8	1	VIP	Si	41	70703244	MARIA ANDR	61	780
15	800080027	C.I EXPOFARO S.A.S. .	ANTIOQUIA	MEDELLIN	Confeccione	No	6	1	VIP	Si	7	120306662	JUAN GUILLE	68	12
16	900170403	C.I. RGA LIMITADA .	NORTE DE SANTAN	CUCUTA	Confeccione	No	4	1	VIP	Si	42	0	CARLOS GIO	8	5
17	5487974	JORGE DURAN JAIMES .	NORTE DE SANTAN	CUCUTA	Confeccione	No	6	1	VIP	Si	8	0	CLAUDIA MA	6	6
18	98555019	ALONSO PALACIO BERMUDEZ	RISARALDA	DOS QUEBRA	Confeccione	No	4	1	VIP	Si	10	0	PERNIA ZULE	18	550
19	88236275	YEIRO PORTILLO QUINTERO	NORTE DE SANTAN	CUCUTA	Confeccione	No	6	1	VIP	Si	13	34003495	CARLOS GIO	31	4
20	57449726	ROSALINA PAEZ MARTINEZ	NORTE DE SANTAN	CUCUTA	Confeccione	No	6	2	Potenciales VIP	No	120	175093100	CLAUDIA MA	4	2
21	18503936	GREGORIO BALEN CASTRO	RISARALDA	DOS QUEBRA	Confeccione	No	6	1	VIP	Si	11	737840	PERNIA ZULE	48	6
22	900267979	COMERCIALIZADORA ACOUSTI	DISTRITO CAPITAL	BOGOTA D.C	Confeccione	No	6	1	VIP	Si	12	0	MARTHA YA	8	4
23	804010866	DOTAHERLOGO S.A. .	SANTANDER	BUCARAMAN	Confeccione	No	3	1	VIP	Si	9	28910354,3	INDIRA HURT	305	12
24	22675273	MAYOLIS PABON CHARRIS	ATLANTICO	BARRANQUI	Comercio M	No	2	1	VIP	Si	23	0	ALBERTO EM	92	421
25	900251809	C.I GREENS INTERNACIONAL	NORTE DE SANTAN	CUCUTA	Comercio M	No	1	1	VIP	Si	4	0	CARLOS GIO	17	16



MANUAL DE CREACIÓN DE CÓDIGOS

PRIMERA EDICIÓN

2012

	MANUAL CREACION DE CODIGOS	Edición 1
Código: CM-M-02 Versión:27.01.2012		Página 2 de 6

INTRODUCCIÓN


Comertex S.A. comercializa una gran variedad de productos, los cuales clasifica dentro de dos U.E.N.'s (MATERIAS PRIMAS Y PRODUCTO TERMINADO).

El alto nivel de ingreso de referencias crea la necesidad de estandarizar el proceso de creación de códigos para cada una de ellas con el fin de hacerlo más eficiente.

En el presente manual se define el árbol de clasificación de productos, los procedimientos a seguir para la creación de los códigos, los requisitos que se deben cumplir para asignar el nombre a cada producto de acuerdo a la línea a la cuál pertenece y la información que se debe ingresar al sistema de información SEVEN.



ELABORADO POR Cargo: Coordinadora de Proyectos	REVISADO POR Cargo: Analista de Auditoria interna	AUTORIZADO POR Cargo: Gerente de Compras
----------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------

	MANUAL CREACION DE CODIGOS	Edición 1
Código: CM-M-02 Versión: 27.01.2012		Página 3 de 6

1. OBJETIVO: Establecer la metodología para crear códigos en la base de datos del sistema de información de Comertex S.A., definiendo la descripción, la referencia, el código de clasificación y los demás elementos que influyen en esta definición para cada uno de los productos que ofrece la compañía.

2. ALCANCE: Este documento rige a partir de su fecha de aprobación y aplica para la creación de los productos en SEVEN ERP (comerciales y administrativos) de todas las líneas de negocio de Comertex S.A. a nivel nacional.

3. DEFINICIONES:

Árbol de clasificación: Agrupación de productos de acuerdo a su uso u origen.

Línea de productos¹: Nivel del árbol que agrupa a los productos según utilización en el mercado.

Categoría de productos²: Nivel del árbol de clasificación donde se definen las características generales de los productos.

Clase de producto: En el nombre, identifica el producto por su característica más general, en la mayoría de los casos, es igual a la categoría del producto.

Atributo de producto: Especifica las características particulares de los productos.

Nombre: Denominación dada al producto para diferenciarlo de los demás de su tipo. Este nombre debe ser claro y es el asignado por el proveedor en la descripción alterna el cual se puede cambiar para su comercialización.

Calidad: Indica la calidad del producto (A, B, C).

Color o Diseño: Indica el color del producto, si éste maneja diseño se debe especificar la respectiva pinta. Cuando el color sea combinado se nombran ambos separado por un (/) sin dejar espacios.

Referencia: Indica el código de la base, acabado, color o diseño y la pinta.

Tipo de producto: Diferencia contablemente los productos comercializados.

¹ Ver Anexo 1: DEFINICIÓN LÍNEA DE PRODUCTOS

² Ver Anexo 2: CATEGORÍA DE PRODUCTOS

ELABORADO POR Cargo: Coordinadora de Proyectos	REVISADO POR Cargo: Analista de Auditoría Interna	AUTORIZADO POR Cargo: Gerente de Compras
----------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------

 Comertex [®] Gente Confiable	MANUAL CREACION DE CODIGOS	Edición 1
Código: CM-M-02 Versión:27.01.2012		Página 4 de 6

4. DESARROLLO:

Para la creación de nuevos códigos en el sistema es necesario definir la siguiente información:

➤ NOMBRE DE PRODUCTOS COMERTEX³

Para la creación del nombre al nuevo producto se debe tener en cuenta la línea a la cual pertenece y dependiendo de ésta se le asignan los diferentes parámetros:

LÍNEA	NOMBRE DEL PRODUCTO						
TEJIDO PLANO	Clase de producto	Nombre	Especificaciones	Color o Diseño/Pinta	Calidad	Referencia	
TEJIDO PUNTO	Clase de producto	Nombre	Color o Diseño/Pinta	Calidad	Referencia		
NO TEJIDO	Clase de producto	Nombre	Dimensión (mm)	Peso (g)	Color	Calidad	Referencia
CALZADO	Clase de producto	Nombre	Color	Talla	Referencia		
HOGAR	Clase de producto	Nombre	Dimensión (mm)	Peso (g)	Acabado	Color	Calidad
VESTUARIO	Clase de producto	Nombre	Color o Diseño	Calidad	Talla	Referencia	

➤ CÓDIGO DE CLASIFICACIÓN

Los códigos de clasificación se asignan, manteniendo coherencia con la estructura definida para el árbol de clasificación de SEVEN (Ver Anexo 4: Árbol de clasificación).

El código de clasificación de productos presenta igual estructura para todas las líneas de negocio, el código está dividido en seis niveles previamente definidos donde se resume toda la información referente al producto (Línea, Categoría...)

³ Ver Anexo 3: NOMBRE DE PRODUCTOS

ELABORADO POR Cargo: Coordinadora de Proyectos	REVISADO POR Cargo: Analista de Auditoría Interna	AUTORIZADO POR Cargo: Gerente de Compras
----------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------

 Comertex <small>gente confiable</small>	MANUAL CREACION DE CODIGOS	Edición 1
		Página 5 de 6
Código: CM-M-02 Versión:27.01.2012		

La definición de cada uno de los niveles es independiente dentro de cada línea de negocio, por el tipo de producto manejado dentro de ésta.

AABBCCDDEEFF	
AA	Representa los dos dígitos asignados a la línea de negocio a la cual pertenece el producto.
BBB	Representa los tres dígitos asignados a los proveedores con los que la empresa mantiene una relación comercial.
CC	Indica la categoría a la cual pertenece el producto de acuerdo a la línea en la cual se encuentra clasificado.
DD	Hace referencia a los dígitos asignados al grupo de los productos según la categoría a la cual pertenece el producto.
EE	Representa los dos dígitos asignados al subgrupo del producto teniendo en cuenta el grupo del producto.
FF	Representa los dos dígitos asignados a la clase del producto. Varía de acuerdo a la UEN del producto.
00	Cuando alguno de los niveles de clasificación no aplique para el producto que se esté clasificando se le asigna el código ÚNICO (00).

➤ MAESTRO DE PRODUCTOS⁴

Para la creación de productos es necesario diligenciar la siguiente información dentro de los diferentes maestros de SEVEN:

Maestro de productos:

- Nombre
- Nombre alternativo
- Tipo de producto comercial

Ver Anexo: CM-IN-03 INSTRUCTIVO CREACIÓN DE CÓDIGOS EN SEVEN

ELABORADO POR Cargo: Coordinadora de Proyectos	REVISADO POR Cargo: Analista de Auditoría Interna	AUTORIZADO POR Cargo: Gerente de Compras
----------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------

	MANUAL CREACION DE CODIGOS	Edición 1
Código: CM-M-02 Versión:27.01.2012		Página 6 de 6

- Subproductos producción (98)
- Productos comerciales con IVA del 16% (100)
- Productos comerciales sin IVA (101)
- Productos comerciales con IVA del 10%
- Compras de mercancía de importaciones (128)
- Productos de confección (servicios) (469)
- Producto terminado confección (493)
- Clasificación del producto (árbol del producto)
- Unidad base SKU
 - Metros (M)
 - Kilogramos (KG)
 - Unidad (UNDE)
 - Pares (PAR)
- Código de barras (si aplica)
- Origen
 - Nacional
 - Importado
- Clase de producto
 - Producto terminado (PT)
 - Materias primas (MP)
 - Servicio (SV)
 - Suministro/Insumo (SU)

Datos de compra:

- Proveedor asignado

Maestro de costos:



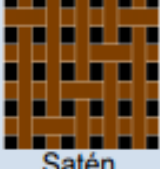




- Instalación
- Conjunto
- Costo del nivel
- Fuente



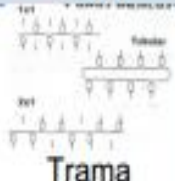


Datos de almacenamiento:

- Bodega por defecto almacenamiento
- Ubicación por defecto
- Bodega por defecto de consumo

ELABORADO POR Cargo: Coordinadora de Proyectos	REVISADO POR Cargo: Analista de Auditoría interna	AUTORIZADO POR Cargo: Gerente de Compras
----------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------

Anexo D. DEFINICIÓN DE LÍNEAS DE PRODUCTO

LÍNEA	CÓDIGO	DEFINICIÓN	IMAGEN
TEJIDO PLANO	01	<p>El tejido plano está conformado por dos tipos de hilado en su estructura: el que va longitudinalmente a la tela (hilado de urdimbre) y el otro es el que va a lo ancho de la tela (hilado de trama). El tejido plano se divide en tres tipos (por el entrelazado de la trama con la urdimbre):</p> <ul style="list-style-type: none"> • Tafetán • Sarga • Raso o satén 	 <p>Tafetán</p>
			 <p>Sarga</p>
			 <p>Satén</p>
NO TEJIDO	02	<p>Los no tejidos son un tipo de tela producidas al formar una red con fibras que se unen por procedimientos mecánicos, térmicos o químicos, pero sin ser tejidas. Es una lámina, velo o napa de fibras flexibles y porosas, sin trama. Para su fabricación las fibras textiles se vuelcan en una bandeja de forma aleatoria sin que predomine ninguna dirección. El material no tejido se produce principalmente de fibras textiles sintéticas como el polipropileno (PP) y el poliéster (PET).</p>	 <p>No tejido</p>
			 <p>Polipropileno</p>
			 <p>Poliéster</p>
HOGAR	03	Incluye: utensilios de cocina, ropa de cama, implementos de baño, decoración, muebles y otros productos para el hogar	 <p>Cocina</p>

		Alcoba
<p>TEJIDO PUNTO 04</p> <p>El tejido de punto es el que se forma con al menos un hilo que se entrelaza consigo mismo. Hay dos variantes fundamentales en las que se divide a los tejidos de punto:</p> <ul style="list-style-type: none"> • Por Trama • Por Urdimbre 		Baño
<p>CALZADO 05</p> <p>Cualquier prenda que sirve para cubrir y resguardar el pie y a veces también la pierna</p>		Trama
<p>VESTUARIO 29</p> <p>La ropa (también llamada vestimenta, atuendo o indumentaria) es el conjunto de prendas generalmente textiles fabricadas con diversos materiales y usadas para vestirse, protegerse del clima adverso y en ocasiones por pudor (ropa interior). En su sentido más amplio, incluye también los guantes para cubrir las manos y gorros, gorras y sombreros para cubrir la cabeza.</p>		Urdimbre
		

Anexo E. CLASIFICACIÓN DE PRODUCTOS POR GRUPO

	LÍNEA	PROVEEDOR		CATEGORIA		GRUPO		SUBGRUPO		CLASE
01	TEJIDO PLANO	PROVEEDOR	01	CONFECCION DOTACIONES	01	INDIGOS	01	MAGNUN	01	RCF
							02	CALIPSO	02	MODA
							03	TRONIC	03	RBR
							04	OTROS		
					02	OXFORD	01	UNICOLOR		
							02	MIL RAYAS		
					03	DRIL LABORAL	01	RAZA		
							02	VULCANO		
							03	WORK		
							04	GOLDEN		
							05	TROYANO		
							06	OTROS		
					04	GABARDINA	01	TEMPO		
							02	OTROS		
					05	LINO	01	ZERMAT		
							02	OTROS		
					06	POPELINA RIGIDA	01	SUPERIOR COLORES		
							02	SUPERIOR BLANCA		
					07	DACRON	01	TEÑIDO		
							02	ESTAMPADO		
							03	APT		
			02	CONFECCION FEMENINA	01	INDIGOS ELASTICOS	01	TEÑIDO		

		02	NATURAL
02	INDIGOS RIGIDOS	01	TEÑIDO
		02	NATURAL
03	TAFETA ELASTICO	01	TEÑIDO
		02	ESTAMPADO
		03	PRETEÑIDO
		04	APT
04	CORDUROY ELASTICO	01	TEÑIDO
		02	ESTAMPADO
		03	APT
05	PEPELINA ELASTICA	01	UNICOLOR
		02	ESTAMPADA
		03	BORDADA
		04	PRETEÑIDO
		05	APT
06	PEPELINA RIGIDA	01	UNICOLOR
		02	ESTAMPADA
		03	BORDADA
		04	PRETEÑIDO
		05	APT
07	TRANSPARENCIAS	01	UNICOLOR
		02	ESTAMPADO
		03	PRETEÑIDO
08	LINO	01	TEÑIDO
		02	ESTAMPADO
09	BLANCURAS	01	LISA
		02	DOBBIES

					03	BORDADA
					01	TEÑIDO
				10	OTROS	02 ESTAMPADO
						03 PRETEÑIDO
						04 BORDADA
				01	INDIGOS RIGIDOS	01 TEÑIDO
						02 NATURAL
				02	C. SEER SUCKER	01 TEÑIDO
						02 PRETEÑIDO
				03	CAMISERO	01 TEÑIDO
						02 PRETEÑIDO
				04	P. TWILL RIGIDO	01 TEÑIDO
						02 ESTAMPADO
						03 PIGMENT DYE
						04 PRETEÑIDO
						05 APT
		03	CONFECCION INFANTIL	05	P. TAFETA RIGIDO	01 TEÑIDO
						02 ESTAMPADO
						03 PIGMENT DYE
						04 PRETEÑIDO
						05 APT
				06	P. TWILL ELASTICO	01 TEÑIDO
						02 ESTAMPADO
						03 PRETEÑIDO
						04 APT
				07	P. CANVAS	01 TEÑIDO
						02 ESTAMPADO

				03	PRETEÑIDO
				04	APT
			08	P. PIQUE	01 TEÑIDO
					02 ESTAMPADO
					03 PRETEÑIDO
					04 APT
			09	ETAMINA	01 TEÑIDO
					02 ESTAMPADA
					03 APT
			10	POPELINA	01 TEÑIDO
					02 ESTAMPADA
					03 APT
			11	LINO	01 TEÑIDO
					02 ESTAMPADA
					03 APT
			12	PIQUE	01 TEÑIDO
					02 ESTAMPADA
					03 APT
			13	CORDUROY	01 TEÑIDO
					02 ESTAMPADO
					03 APT
			14	OTROS	01 TEÑIDO
					02 ESTAMPADO
					03 APT
		04	CONFECCION MASCULINA		01 TEÑIDO
			01	INDIGOS RIGIDOS	02 NATURAL
			02	P. TWILL RIGIDO	01 TEÑIDO

				02	ESTAMPADO
				03	PRETEÑIDO
				04	APT
		03	P. CANVAS	01	TEÑIDO
				02	ESTAMPADO
				03	PRETEÑIDO
				04	APT
		04	P. TAFETA RIGIDO	01	TEÑIDO
				02	ESTAMPADO
				03	PRETEÑIDO
				04	APT
		05	CORDUROY RIGIDO	01	TEÑIDO
				02	APT
		06	TELA DE BOLSILLO	01	CRUDA
				02	TENIDA
				03	ESTAMPADA
		07	CAMISERO	01	TEÑIDO
				02	PRETEÑIDO
				03	CEDRO
				04	CARACOLI
				05	MADRID
		08	DACRON	01	TEÑIDO
				02	APT
	05	CONFECCION ROPA QUIRURGICA	00	UNICO	00 UNICO
	06	CONFECCION ROPA DEPORTIVA	01	IMPERMEABLE	01 TEÑIDA
					02 ESTAMPADA

				01	GENERO	01	BLANCO	
						02	COLORES	
						03	ESTAMPADO	
				02	LONA	01	CRUDA	
						02	ESTAMPADA	
				03	TELA TOALLA	01	BLANCO	
						02	COLORES	
						03	ESTAMPADO	
				04	BAYETILLA	01	BLANCO	
						02	ROJO	
						03	COLORES	
			07	INDUSTRIAL	05	LIMPIATODO	01	TEÑIDO
						02	ESTAMPADO	
				06	TELA PAÑAL	01	BLANCO	
						02	COLORES	
						03	ESTAMPADO	
				07	VELO SUIZO	01	TEÑIDO	
						02	ESTAMPADO	
						03	BORDADO	
				08	SATIN	01	TEÑIDO	
						02	ESTAMPADO	
				09	OTROS	01	TEÑIDO	
						02	ESTAMPADO	
		08	PUBLICIDAD	00	UNICO	00	UNICO	

Anexo F. MODELOS DE REDES BAYESIANAS

The screenshot displays the IBM SPSS Modeler interface. The main window, titled 'SEGMENTO', shows a Bayesian network diagram on the left and a bar chart titled 'Importancia del predictor' on the right. The network diagram has four nodes: 'DEPARTAMENTO' (blue circle, predictor), 'Algodón' (blue circle, predictor), 'VIP' (blue circle, predictor), and 'SEGMENTO' (red circle, objective). Arrows indicate dependencies: 'DEPARTAMENTO' and 'Algodón' are parents of 'VIP', and 'SEGMENTO' is a parent of 'Algodón'. The bar chart shows the importance of each predictor for the objective 'SEGMENTO'. The x-axis ranges from 0.0 to 1.0, with 'Menor importancia' on the left and 'Mayor importancia' on the right. The bars show that 'DEPARTAMENTO' has the highest importance (approximately 0.8), followed by 'VIP' (approximately 0.2), and 'Algodón' (approximately 0.1).

Predictor	Importancia
DEPARTAMENTO	0.8
VIP	0.2
Algodón	0.1

Red clientes 2* - IBM® SPSS® Modeler

Archivo Edición Insertar Ver Herramientas Supernodo Ventana Ayuda

SEGMENTO

Modelo Configuración Resumen Anotaciones

Archivo Generar Ver Presentación preliminar

SEGMENTO

EXCEL

BBdd Clientes (ok).x.

SEGMENTO

SEGMENTO

Favoritos Orígenes

SQL Archivo var. Prepar

Base de datos

Ver: Básico Restablecer

Aceptar Cancelar

Ver(W): Importancia del predictor

Aplicar Restablecer

277MB / 396MB

Resultados Modelos

DM Clases

proyecto no guardado)

Comprensión del negocio

Comprensión de los datos

Preparación de los datos

Modelado

Evaluación

Distribución

plano Base de datos

Red bayesiana

Tipo

- Predictores
- Objetivo

Importancia

- 1,0
- 0,8
- 0,6
- 0,4
- 0,2
- 0,0

DEPARTAMENTO

Algodón

SEGMENTO

VIP

Importancia del predictor

Objetivo: SEGMENTO

DEPARTAMENTO

Algodón

VIP

Menor importancia Mayor importancia

Red clientes 2* - IBM® SPSS® Modeler

Archivo Edición Insertar Ver Herramientas Supernodo Ventana Ayuda

SEGMENTO

Modelo Configuración Resumen Anotaciones

Archivo Generar Ver Presentación preliminar

Red bayesiana

Tipo

- Predictores (Azul)
- Objetivo (Rojo)

Importancia

- 1,0
- 0,8
- 0,6
- 0,4
- 0,2
- 0,0

DEPARTAMENTO

Algodón

SEGMENTO

VIP

Importancia del predictor

Objetivo: SEGMENTO

Predictor	Importancia
DEPARTAMENTO	0,8
VIP	0,2
Algodón	0,1

Ver: Básico Restablecer

Aceptar Cancelar

Ver(W): Importancia del predictor

Aplicar Restablecer

Resultados Modelos

Clases

proyecto no guardado)

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Distribución

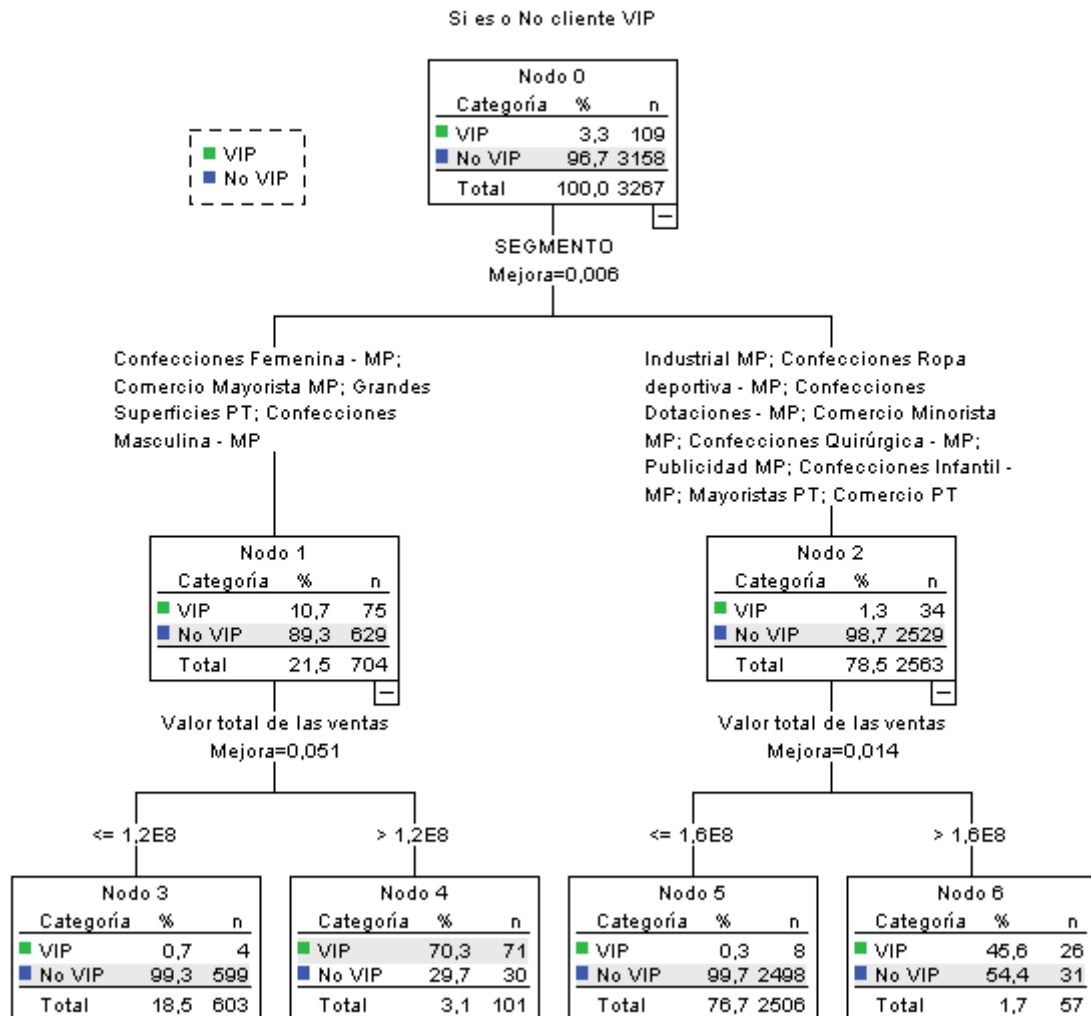
Base de datos Archivo var. Prepar

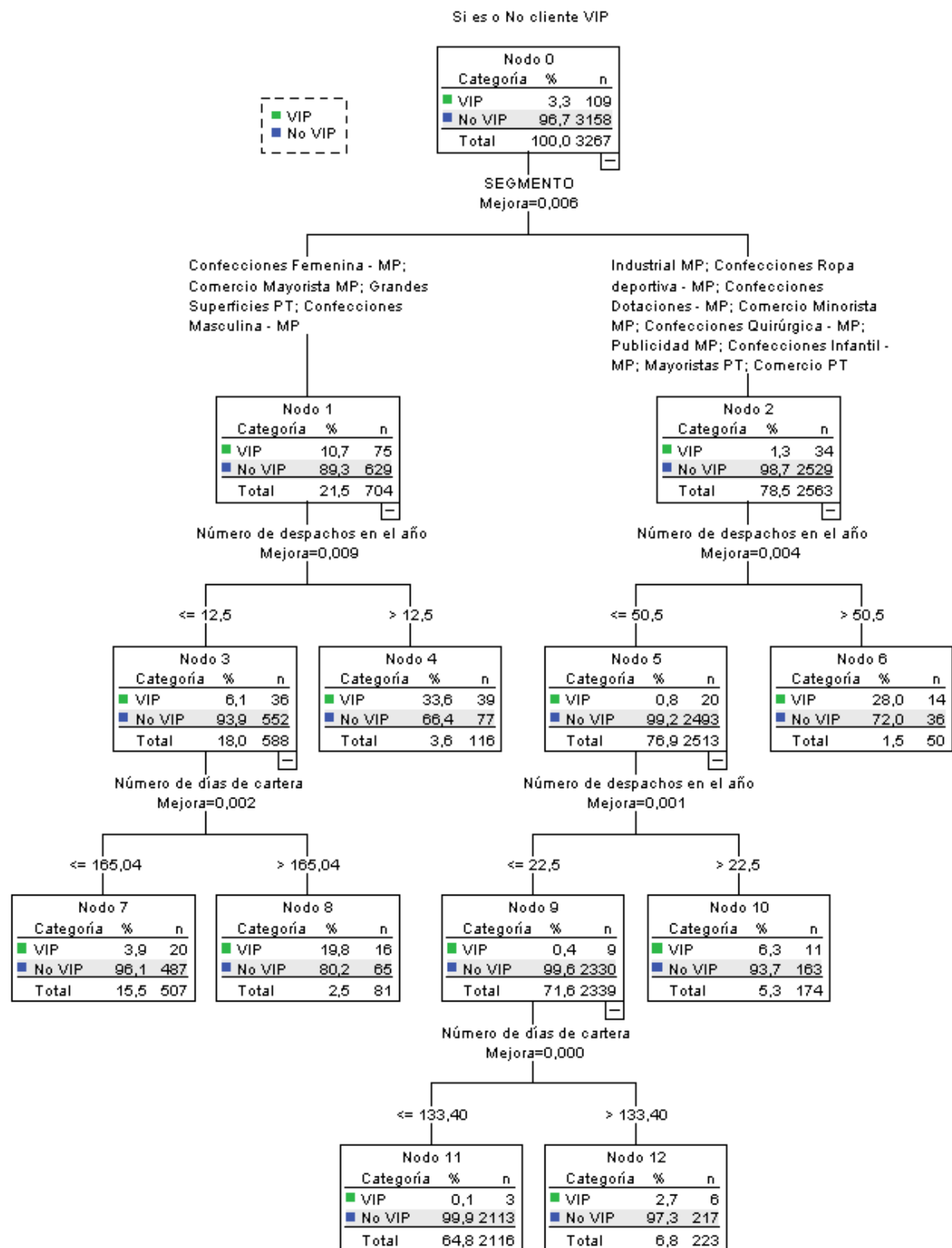
SQL

plano Base de datos

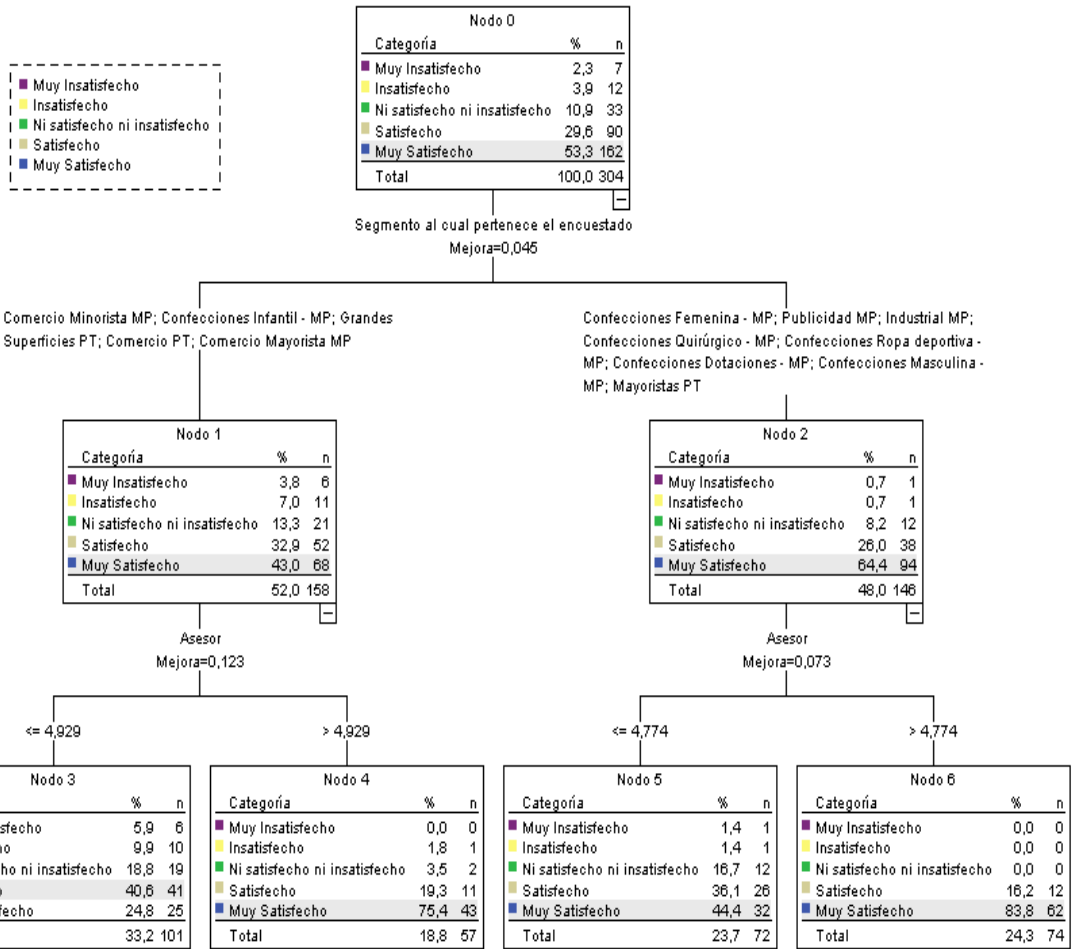
288MB / 412MB

Anexo G. CLASIFICACIÓN DE CLIENTES MEDIANTE CART





Satisfacción



Anexo H. CARACTERIZACIÓN DE SEGMENTOS

Criterios de evaluación	C. May MP	C. Min MP	Confecciones					
			Dotaciones	Femenina	Infantil	Masculina	Quirúrgica	Ropa deportiva
N. de clientes	120	647	431	262	225	305	110	114
Ventas (%participación)	9,1%	6,4%	8,0%	11,5%	2,7%	12,1%	1,3%	2,0%
Líneas producto (representativas)	Tejido Plano (82%)	No tejidos (95%) Tejido Plano Tejido Punto	Tejido Plano(82%)	Tejido Plano (94%)	Tejido Plano Tejido Punto	Tejido Plano(94%)	No Tejidos	Tejido Punto(90%)
Grupo de producto (representativos)	Indigos	Gran número de grupos	Driles Indigo	Indigos	Gran número de grupos	Indigos	Bonlam	Elastomeros Deportivos
Devoluciones (% participación)	3,5%	3,1%	3,6%	4,6%	1,4%	8,8%	0,2%	1,4%
Cartera (Días para el pago)	89	71	82	110	109	117	48	62
Pedidos	Cada 10 días Valor: 3,5 millones	Cada 32 días Valor: 1,5 millones	Cada 31 días Valor: 2,6 millones	Cada 63 días Valor: 1,8 millones	Cada 37 días Valor: 2 millones	Cada 53 días Valor: 9,7 millones	Cada 37 días Valor: 2 millones	Cada 28 días Valor: 2,2 millones
Utilidad Operacional (% participación)	1,7%	-6,6%	4,7%	18,7%	3,2%	17,5%	2,9%	-3,0%
Número de encuestados	8	61	41	23	12	37	4	3
Satisfacción	62,5%	85,2%	90,2%	82,6%	81,8%	89,5%	100,0%	100,0%
Aspectos importantes a la hora de la compra	Precio Calidad del producto	Calidad del producto Precio	Calidad del producto Precio	Calidad del producto Precio	Precio Calidad del producto	Calidad del producto Precio	Calidad del producto Precio	Calidad del producto Precio Tiempo de entrega
Razones de compra	Producción Distribución	Comercialización Confección	Confecciones	Confección Precio	Confección	Confección Comercialización	Confección	Confección
Razones de uso	Confección prendas de vestir, venta	Ropa dama, almohadas, Toallas, Zapatos	Uniformes, Dotaciones, Ropa deportiva	Jeans, terminación de prendas, Comercialización	Ropa para bebe, pijamas	Pantalones, camisas	Ropa hospitalaria	Ropa deportiva y uniformes
Busqueda de información	R. de ventas Internet	R. de ventas Internet Distribuidores	Internet Distribuidores Terceros	Distribuidores Internet	R. de ventas Internet	Distribuidores R. de ventas	Distribuidores R. de ventas	Internet Distribuidores
Entregas	Actual= 5 días Ideal= 2 días	Actual= 4 días Ideal= 2 días	Actual= 3 días Ideal= 2 días	Actual= 5 días Ideal= 3 días	Actual= 2 días Ideal= 3 días	Actual= 4 días Ideal= 3 días	Actual= 5 días Ideal= 4 días	Actual= 1 días Ideal= 1 días
Inventario	No les gusta	5 días	2 días	8 días	2 días	2 días	45 días	No les gusta

Criterios de evaluación	Industrial	Publicidad	Producto Terminado		
			Grandes Superficies	Mayoristas	Comercio
N. de clientes	280	94	17	153	509
Ventas (%participación)	40,4%	0,6%	1,6%	2,4%	1,9%
Líneas producto (representativas)	Algodón No Tejidos Tejido Plano	No Tejidos (95%) Tejido Punto Tejido Plano	Hogar (97,5) Calzado	Calzado (95,7) Hogar	Hogar (80%) Calzado
Grupo de producto (representativos)	Bonlam Importadas	Bonlam Punto sencillo	Baño Infantil Descanso	Infantil Calzado deportivo Deportivo	Baño Alcoba Infantil
Devoluciones (% participación)	66,2%	0,2%	2,0%	2,5%	2,5%
Cartera (Días para el pago)	62	43	114	88	78
Pedidos	Cada 57 días Valor: 2,7 millones	Cada 63 días Valor: 1,9 millones	Cada 2 días Valor: 547 millones	Cada 29 días Valor: 2,2 millones	Cada 103 días Valor: 1,8 millones
Utilidad Operacional (% participación)	75,5%	-0,5%	0,9%	-17,2%	2,2%
Número de encuestados	28	3	3	8	74
Satisfacción	92,9%	100,0%	100,0%	87,5%	68,9%
Aspectos importantes a la hora de la compra	Calidad del producto Precio Tiempo de entrega	Calidad del producto Precio	Calidad del producto Diseño del producto Tiempo de entrega Asesoría comercial Servicio pos venta	Calidad del producto Precio	Calidad del producto Diseño del producto Precio
Razones de compra	Confección de gran de volumen	Producción, venta	Calidad Precio	Precio Rapidez para hacer negocios	El cliente lo pide Precio
Razones de uso	Almohadas, lencería, cubrelechos, cojines	Camisetas, ropa desechable y publicidad	Comercialización	Comercialización	Comercialización
Busqueda de información	Terceros Internet	Internet Distribuidores	R. de ventas Internet	R. de ventas Internet Revistas especializadas Eventos de la compañía	R. de ventas Internet
Entregas	Actual= 4 días Ideal= 3 días	Actual= 2 días Ideal= 1 días	Actual= 9 días Ideal= 5 días	Actual= 4 días Ideal= 3 días	Actual= 7 días Ideal= 3 días
Inventario	13 días	1 días	2 días	4 días	7 días

Anexo I. PROMESA DE SERVICIO

DEBILIDADES	FORTALEZAS	RECOMENDACIONES
<p>COMERTEX S.A. presenta algunas limitaciones en aspectos relacionados con el servicio como sigue:</p> <ul style="list-style-type: none"> × Asesor comercial: el 10% de los clientes NO está satisfecho con la asesoría que se le brinda por parte de éste. Las principales causas son: <ul style="list-style-type: none"> ○ El asesor no lo visita ○ El asesor no sabe vender × Producto: Casi el 20% de los clientes muestra inconformidad con el producto debido a: <ul style="list-style-type: none"> ○ Disminución de la calidad del producto ○ No hay variedad de productos × Entregas: 10% de los clientes muestra insatisfacción con la entrega de los pedidos debido a: <ul style="list-style-type: none"> ○ Demora en la entrega del pedido ○ No envían lo que el cliente pide 	<p>Según los datos de la encuesta, COMERTEX S.A. cuenta con un buen nivel de servicio.</p> <p>COMERTEX S.A. cuenta con algunas herramientas que pueden facilitar tanto la mejora en el nivel de servicio como la relación con los clientes. Por ejemplo, el manejo del ERP debe ser al máximo, existe la percepción de que el uso que se le da actualmente es muy básico.</p> <p>COMERTEX S.A. cuenta con datos históricos, por lo tanto, es fundamental que esos datos se conviertan en información que permita optimizar la toma de decisiones.</p> <p>COMERTEX S.A. cuenta con los datos para realizar el análisis de la canasta por segmento</p>	<p>Soportados en los datos se generan una serie de recomendaciones que podrían ser el camino a seguir; se parte desde acciones que pueden ser efectuadas en poco tiempo, pasando por acciones más elaboradas y llegando al deber ser.</p> <ol style="list-style-type: none"> 1. Creación de un nivel mínimo de servicio: COMERTEX S.A. creará una política de servicio que incluya los siguientes aspectos <ul style="list-style-type: none"> ✓ Asesor comercial ✓ Productos ✓ Entregas ✓ Atención quejas y reclamos ✓ Información <p>Adicional, a la encuesta actual se deben adicionar preguntas que se encuentran posterior a esta tabla resumen.</p> 2. A partir de la construcción del nivel mínimo de servicio COMERTEX S.A. puede generar valor a través del nivel de servicio diferenciado para

<p>× Atención telefónica: el 16% de los clientes evidencia molestia con el servicio de la atención telefónica por las razones que siguen:</p> <ul style="list-style-type: none"> ○ No contestan las llamadas ○ No dan soluciones rápidas por teléfono <p>× Presencia en ferias o eventos: la participación de los clientes en este aspecto es muy baja, solo el 17% manifestó su presencia en estos eventos e informan que éstos los motivan a la compra.</p> <p>× Devoluciones: representan el 4% de las ventas y el 85% está concentrado en 3 criterios sobre los que se podría generar acciones inmediatas.</p> <p>Adicional al nivel de servicio, existen otros aspectos en los cuales COMERTEX S.A. debe trabajar, entre otros:</p> <ul style="list-style-type: none"> ✓ Recolección y tratamiento de datos ✓ Sacar el máximo provecho de las tecnologías de información 		<p>cada segmento. En el capítulo 5.3 se da un punto de partida para la diferenciación del servicio de acuerdo a los datos de la encuesta y otros. En este punto es conveniente evaluar la posibilidad de crear una estrategia para los clientes VIP, algunas ideas se encuentran al final de este anexo.</p> <p>3. Elevar el nivel de servicio</p> <ul style="list-style-type: none"> ✓ Gestión de devoluciones ✓ Gestión de reclamaciones ✓ Análisis de venta cruzada ✓ Clientes más propensos a responder a ofertas y promociones ✓ Predicciones de venta <p>4. Nivel de servicio y marketing online: desarrollar un ambiente en línea, puede ser a través de la página web de la empresa o mediante una red social, que permita conocer de manera oportuna las necesidades de los clientes en cuanto a servicio y marketing.</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Preguntas que pueden ser adicionadas en la encuesta de servicio

- Cómo prefiere el empaque de los pedidos.
- Qué día(s) de la semana acostumbra a solicitar el pedido.
- Qué día(s) de la semana prefiere recibir el pedido.
- Cuántas visitas al mes considera necesarias por parte del asesor comercial (número de veces).
- Qué otro medio de comunicación le gustaría tener con el asesor (telefónico-correo electrónico-etc.)
- Qué tipo de servicio post-venta requiere (asesoría sobre tendencias en las confecciones, diseños, etc.)
- Forma o medio de pago (Al asesor, consignación, transferencia bancaria, etc.)
- Cuántos días de crédito le gustaría recibir por parte de COMERTEX S.A.
- Cómo les parecería el servicio de quejas y reclamos a través de la página web.
- Cada cuánto realizan un pedido (número de días).
- Opinión acerca de la construcción de una aplicación para teléfonos inteligentes que les permita generar sus quejas y reclamos de manera inmediata.
- Opinión acerca de la página web de COMERTEX S.A.
- Evaluar las expectativas versus las percepciones que tiene el cliente sobre el nivel de servicio.
- Que tiempo considera necesario para que el asesor atienda sus necesidades (número o cantidad de horas).

✓ **CLIENTES VIP**

Como parte de la estrategia de servicio diferenciada se puede crear una estrategia para los clientes VIP, estrategia que puede incluir aspectos como:

- Invitación a eventos especiales
- Servicios gratuitos
- Ser los primeros en recibir ofertas y promociones

- Atención personalizada (por ejemplo por los gerentes directamente)
- Acreditación de cliente VIP