



**ESCUELA SUPERIOR POLITÉCNICA  
DEL LITORAL**

**Instituto de Ciencias Matemáticas**

**Ingeniería en Estadística Informática**

**“Modelo de Oportunidad de Mejora para el  
cliente en la elección de su plan celular  
mediante el diseño de una datawarehouse y  
árboles de decisión”**

**TESIS DE GRADO**

Previa a la obtención del Título de:

**INGENIERA EN ESTADÍSTICA INFORMÁTICA**

Presentado por:

**Lorena Patricia Calvas Bayas  
Ángela Elvira Vásconez Vásconez**

**GUAYAQUIL – ECUADOR**

**2010**

# AGRADECIMIENTO

Ángela...

A Dios, mis padres, mi esposo y a cada docente que a lo largo del camino de mi vida estudiantil puso su granito de arena para hacer de mí una profesional y un ser humano perfectamente normal.

Lorena...

A Dios, a mi familia y a todas las personas que han hecho posible que pueda cumplir con este gran anhelo, el poder obtener mi título de ingeniera.

# DEDICATORIA

Ángela...

A Dios y a las personas más importantes en mi vida, mis padres, mi esposo, mi suegra, mis hijas y hermanos, pues son el pilar de mi vida y razón de existir, los amo.

Lorena...

Para mis amores Xavier y Samuel por ser el gran pilar de mi vida, a mi abuelita por ser la mujer que ilumina mi camino, a mis padres por empujarme en mis estudios y a mi hermana por darme sus fuerzas para seguir adelante.

# **Tribunal de Graduación**

---

Mat. Jhon Ramírez  
Delegado

---

Msc. Pedro Fabricio Echeverría Briones  
Director de Tesina

# Declaración Expresa

“La responsabilidad del contenido de este trabajo de graduación nos corresponde exclusivamente; y el patrimonio intelectual de la misma a la Escuela Superior Politécnica del Litoral”

(Reglamento de graduación de la ESPOL)

---

Lorena Patricia Calvas Bayas

---

Ángela Elvira Vásconez Vásconez

# RESUMEN

Para fines de este proyecto hemos denominado a la empresa de telefonía celular como “Cellcorpo”.

Cellcorpo es una empresa de telefonía celular con diversos tipos de productos de voz, de los cuales se han escogido dos, es decir, los que tienen más captación en el mercado, estos son: Plan Controlado y Plan Pre-Pago.

Producto Controlado es aquel que tiene fijo una cantidad de minutos y/o mensajes, dependiendo del que haya escogido el cliente al momento de la compra de su equipo celular. Existe una gama de planes para este producto, los cuales pueden ser escogidos por el cliente ya sea esta persona natural o jurídica.

Producto Pre-Pago es aquel que tiene como característica ofrecer al cliente tarjetas prepago en donde se activan directamente y dependiendo del monto existe un tiempo de vigencia de dicha tarjeta.

Realizamos un estudio con árboles de decisión con la base de datos de la empresa para obtener como resultado a los clientes que debemos escoger para presentarles una mejor opción de acuerdo a sus hábitos de consumo durante los últimos tres meses de utilización de su producto celular.

# ÍNDICE GENERAL

	Pág.
RESUMEN	II
ÍNDICE GENERAL	III
ABREVIATURAS	IV
SIMBOLOGÍA	V
ÍNDICE DE FIGURAS	VI
ÍNDICE DE GRÁFICOS	VII
ÍNDICE DE TABLAS	VIII
INTRODUCCIÓN	XV

## I. EMPRESA DE TELEFONÍA CELULAR

1.1 Breve Reseña.....	1
1.1.1 Antecedentes.....	1
1.1.2 Descripción del Entorno.....	2
1.1.3 Definición del Problema.....	4
1.1.4 Solución del Problema.....	4

## II. DISEÑO DEL DATAWAREHOUSE

2.1 Introducción.....	6
2.2 Diagrama de Influencia.....	6
2.3 Modelo de minería de datos.....	8
2.3.1 Datamining (Minería de datos).....	8
2.3.2 Datamart.....	9

2.3.3	Datawarehouse.....	10
2.3.4	Tablas de Diseño.....	14
2.3.5	Modelo Multidimensional.....	22

### **III. MODELAMIENTO DE ÁRBOLES DE DECISIÓN**

3.1	Introducción.....	25
3.2	Moldeamiento del Árbol.....	25
3.2.1	Árbol de Decisión.....	26
3.3	Solución del Problema.....	39

### **IV. CONCLUSIONES Y RECOMENDACIONES**

### **BIBLIOGRAFÍA**



## ABREVIATURAS

PPA	Pre-Pago
AUT	Autocontrol
SQL	Structured Query Language
MDX	MultiDimensional eXpressions
ETL	Extracción Transformación y Carga
PROD	Producto
CHAID	Chi-Squared Automatic Interaction Detection
TARIFA_B	Tarifa Básica
TOTAL_TA	Total Tarifa
N	Natural
J	Jurídica
RUC	Registro Único de Contribuyentes
ODBC	Object Data Base Conection
DSS	Sistema de Soporte a la Decisión
EIS	Sistemas de Información Ejecutiva

# SIMBOLOGÍA

$\chi^2$	Ji Cuadrada
$d_k$	K-ésimo Alternativa
$S_k$	K-ésimo Evento Fortuito
\$	Dólares
&	Y
$\leq$	Menor o igual
$>$	Mayor
%	Porcentaje
n	Número de clientes

# ÍNDICE DE FIGURAS

	Pág.
Figura 2.3.1.1	Ilustración de la Carga de trabajo en las fases de un proyecto de Datamining.....9
Figura 2.3.2.1	Ilustración de un Datamart.....10
Figura 2.3.3.1	Ilustración de un Data Warehouse.....12
Figura 2.3.3.2	Ilustración de un Conjunto de Datamarts que conforman un Data Warehouse Corporativo.....15
Figura 2.2.4.1	Tabla Cliente.....16
Figura 2.2.4.2	Tabla Contrato_Cliente.....17
Figura 2.2.4.3	Tabla Numero_Celular.....18
Figura 2.2.4.4	Tabla Detalle_Plan.....19
Figura 2.2.4.5	Tabla Fecha.....20
Figura 2.2.4.6	Tabla Consumo_Cliente_Tarjeta.....21
Figura 2.2.4.7	Tabla Consumo_Cliente_Pla.....22
Figura 2.2.4.8	Tabla Maestro.....23
Figura 3.1	Representación Horizontal de un Árbol de Decisión.....29
Figura 3.2	Representación Vertical de un Árbol de Decisión.....30

# ÍNDICE DE GRÁFICOS

	Pág.
Gráfico 2.1	Diagrama de Influencia..... 7
Gráfico 2.2	Diagrama Estrella del Modelo.....24
Gráfico 2.3	Cubo de Datos.....25
Gráfico 3.1	Árbol de Decisión Predictor: Cantidad.....36

## ÍNDICE DE TABLAS

		Pág.
Tabla I	Análisis de segmentación de las variables predictoras.....	34
Tabla II	Predictoras de la variable producto.....	35

# INTRODUCCIÓN

Para realizar este estudio tomamos la base de datos de la empresa CELLCORPO la cual contiene doscientos veinte y tres mil doscientos cincuenta y seis registros, que es el total de nuestra clientela, quienes poseen dos tipos de planes de telefonía celular, plan prepago o plan controlado.

Tenemos como finalidad determinar un mejor beneficio a los clientes de una empresa de telefonía celular en cuanto al plan que necesite y le sea más conveniente, tomando en cuenta sus hábitos de consumo de los tres últimos meses de utilización de su línea telefónica celular.

Realizamos un estudio con árboles de decisión para obtener como resultado a los clientes que debemos escoger para presentarles una mejor opción con respecto a su plan de telefonía celular actual.

Con la bases de datos de nuestros clientes, realizamos un modelo de minería de datos donde veremos el diagrama estrella con las tablas clientes, consumo cliente plan, consumo cliente tarjeta, contrato cliente, detalle plan, fecha, número celular y la tabla de hecho denominada Maestro, la unión de dichas tablas forman el cubo de datos del modelo.

Con la tabla Maestro se realiza el modelamiento de árboles de decisión, para esto nos valemos de un software denominado AnswerTree que utiliza cuatro tipo de algoritmos, Chaid, Chaid exhaustivo, Arboles de clasificación y regresión (C&RT) y Quest.

**AnswerTree** generó el árbol de decisión mediante el algoritmo CHAID; este modelo es aplicado a la base de datos con el fin de generar listados de clientes susceptibles a gastar más dinero con su actual plan de telefonía celular, por medio de una variable predictora.

Cabe recalcar que para facilidad del estudio este documento se ha dividido en cuatro capítulos: Empresa de telefonía celular, Diseño del Datawarehouse, Modelamiento del Árbol de decisión y; Conclusiones y Recomendaciones.

# **CAPÍTULO I**

## **1. EMPRESA DE TELEFONÍA CELULAR**

### **1.1 Breve reseña**

#### **1.1.1 Antecedentes**

Cellcorpo es una empresa de telefonía celular que brinda diversos tipos de productos a sus clientes, dado que estos brindan promociones, se ha observado un medio de captación del mercado, es por esto que se ve en la necesidad de realizar un análisis para la toma de decisión en los dos tipos de productos considerados los de mayor captación en la empresa.

Se ha observado que ciertos clientes podrían obtener mayor beneficio realizando un cambio en el tipo de producto escogido.



Cellcorpo tiene como objetivo la captación de clientes en dos tipos de productos de llamada celular, Planes Controlados y Plan Prepago, en donde se le pueda ofrecer a cada cliente una mejor opción contra el producto que posee. Para esto se realizará un análisis en los clientes activos sobre los últimos tres meses de consumo del cliente y así poder presentarle opciones que le permitan obtener un mejor beneficio según sus hábitos de consumo, es decir, ver si sus consumos son mayores al plan que tiene actualmente, se le puede brindar un plan que cubra con las expectativas del cliente y con esto atraerlo con un buen servicio y a su vez aumentar el mercado.

La empresa celular basándose en la información que ha recogido en sus bases de datos, deberá tomar la mejor decisión con respecto a dichos planes, para así presentarles a sus clientes una mejor opción mostrándole mayor beneficio al obtener el nuevo plan.

### **1.1.2 DESCRIPCIÓN DEL ENTORNO**

Para el análisis de la toma de decisión del proyecto describimos los productos en consideración, que son prepago y controlado.

A continuación citamos la descripción de cada plan:

#### **PREPAGO**

Este producto tiene como característica ofrecer al cliente tarjetas prepago físicas o virtuales en donde se activan directamente a la línea y dependiendo del monto ingresado posee un tiempo de vigencia.

### **Tiempos de vigencia:**

**\$3:** 7 días - **\$6:** 30 días - **\$10:** 30 días - **\$20:** 45 días - **\$30:** 45 días

### **PLANES CONTROLADOS**

Este producto tiene como característica darle al cliente un monto (el que el elija) para que este sea consumido en un ciclo de facturación, dicho ciclo es el tiempo de consumo que el cliente realiza en cada plan (1 mes de vigencia). Existen una gama de planes para este producto, pueden ser planes personales o empresariales.

Se realizó un estudio en el departamento de Centro de atención al cliente en donde se verifica lo siguiente:

#### **Para planes prepago PPA**

Existen clientes que consumen en promedio \$30.00 por mes, en donde el costo por minuto de este producto es más alto que el de un plan controlado y adicionalmente cada vez que termine el consumo de la tarjeta PPA debe de comprar otra y realizar su respectivo ingreso.

#### **Para planes controlados**

Existen clientes que teniendo un plan controlado consumen más del 40% del valor de su plan, teniendo otras opciones que pueden beneficiar a este.

La empresa maneja diferentes productos de la compañía sin embargo lo que se espera es poder identificar la concentración de usuarios

(productos de voz) y a su vez verificar en donde existen debilidades para encontrar alternativas en mejora de la preferencia de ese producto.

### **1.1.3 DEFINICIÓN DEL PROBLEMA**

Debido a que se ha observado que los clientes tanto en planes prepago como en planes controlados sacarían mayor beneficios con respecto a sus hábitos de consumo celular en los últimos tres meses se debe realizar una toma de decisión para cambiar a un nuevo plan para obtener una mejor ventaja en dicho plan respecto al anterior y así captar mayor cantidad de clientes en el mercado.

La competencia por temas de portabilidad ofrece también una gama de servicios para los planes prepago ocasionando incertidumbre en el cliente la cual ya se está embistiendo con diferentes promociones sin embargo mas allá de eso lo que se busca es poder retener al cliente sin necesidad al cambio.

Los clientes que se encuentran en el plan prepago, podrían obtener mejor beneficio en un plan controlado y clientes que se encuentran en planes controlados podrían cambiarse a otro tipo de plan controlado dependiendo de sus necesidades con el objetivo de abaratar costos y pagando lo necesario por sus minutos o mensajes consumidos.

### **1.1.4 SOLUCIÓN DEL PROBLEMA**

La información se obtendrá de las bases de datos de la empresa de telefonía celular donde se podrá obtener los consumos generados por los clientes de los productos prepago y autocontrol.

Lo que se busca del análisis es poder identificar cuáles son los planes menos solicitados, obtener información de cuáles son los clientes que poseen los planes y poder darle al área comercial en donde existen concentraciones y no concentraciones de mercado para que puedan ver opciones de servicios a brindar a los clientes.

Presentaremos un modelo de toma de decisiones para determinar un mejor beneficio a clientes de la empresa de telefonía celular Cellcorpo en cuanto a planes que necesite y le sea más factible, tanto en su consumo como en su economía.

# **CAPÍTULO II**

## **2. DISEÑO DEL DATAWAREHOUSE**

### **2.1 Introducción**

Este capítulo presenta el análisis de la base de datos de la empresa, el modelo de minería de datos en el cual encontramos los conceptos utilizados como minería de datos, datamart, datawarehouse, las tablas de diseño y el modelo multidimensional. En el modelo de minería de datos encontramos la descripción de cada tabla del diseño que relacionadas junto a una tabla de hecho forman el Modelo Estrella, que da partida al modelo multidimensional.

### **2.2 DIAGRAMA DE INFLUENCIA**

El diagrama de influencia es una herramienta gráfica que muestra las relaciones entre las decisiones, los eventos al azar y las consecuencias para un problema de decisión. Los nodos en un diagrama de influencia se usan para representar las decisiones, los eventos fortuitos y las consecuencias. Se emplea rectángulos o cuadrados para describir los nodos de decisión, círculos u óvalos para describir los nodos fortuitos y rombos para describir los nodos de consecuencia. Las líneas que conectan los nodos, llamadas arcos, muestran la dirección de la influencia que tiene los nodos entre sí.

A continuación representamos el gráfico del Diagrama de Influencia del modelo de datos.

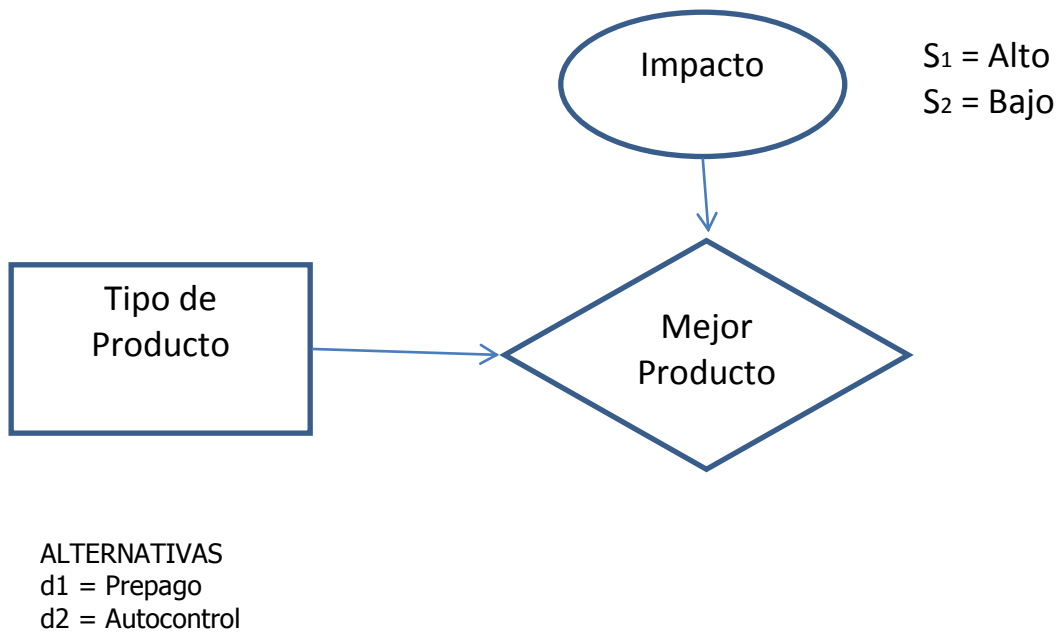


Gráfico 2.1 Diagrama de Influencia

## 2.3 MODELO DE MINERÍA DE DATOS

### 2.3.1 DATAMINING (MINERÍA DE DATOS)

El **datamining** (*minería de datos*), es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Básicamente, el datamining surge para intentar ayudar a comprender el contenido de un repositorio de datos.

Aunque en datamining cada caso concreto puede ser radicalmente distinto al anterior, el proceso común a todos ellos se suele componer de cuatro etapas principales:

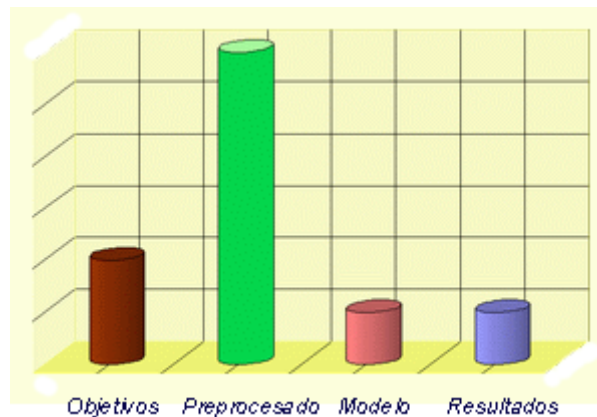
**Determinación de los objetivos.** Trata de la delimitación de los objetivos que el cliente desea bajo la orientación del especialista en datamining.

**Pre procesamiento de los datos.** Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume generalmente alrededor del setenta por ciento del tiempo total de un proyecto de datamining.

**Determinación del modelo.** Se comienza realizando unos análisis estadísticos de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.

**Análisis de los resultados.** Verifica si los resultados obtenidos son coherentes y los coteja con los obtenidos por los análisis estadísticos y de visualización gráfica. El cliente determina si son novedosos y si le aportan un nuevo conocimiento que le permita considerar sus decisiones.





*Figura 2.3.1.1 Ilustración de la Carga de trabajo en las fases de un proyecto de datamining*

En resumen, el datamining se presenta como una tecnología emergente, con varias ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y las personas de negocios; por otro, ahorra grandes cantidades de dinero a una empresa y abre nuevas oportunidades de negocios. Además, no hay duda de que trabajar con esta tecnología implica cuidar un sinnúmero de detalles debido a que el producto final involucra "toma de decisiones". (Tomado de [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cuba.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cuba.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=_2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1))

### **2.3.2 DATAMART**

Un **Datamart** es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la **estructura óptima de datos** para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un datamart puede ser alimentado desde los datos de un **datawarehouse**, o integrar por sí mismo un compendio de distintas fuentes de información. (Tomado de [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cuba.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cuba.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=_2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1))



*Figura 2.3.2.1 Ilustración de un Datamart*

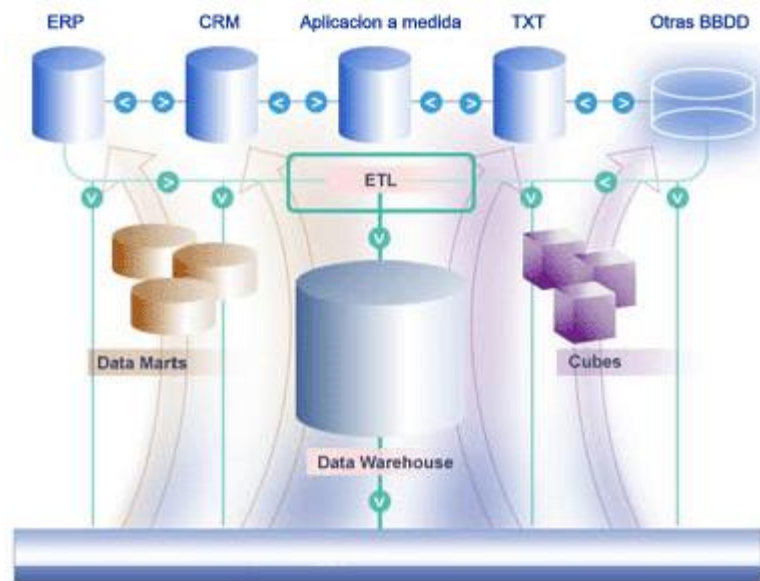
Los datamarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos

### 2.3.3 DATAWAREHOUSE

Un **Datawarehouse** es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. La creación de un datawarehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales, etc.). Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales). (Tomado de [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cuba.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cuba.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=_2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1))



*Figura 2.3.3.1 Ilustración de un Data Warehouse*

Un datawarehouse se caracteriza por ser:

**Integrado:** los datos almacenados en el datawarehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

**Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del datawarehouse. De esta forma, las peticiones de información sobre

clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.

**Histórico:** el tiempo es parte implícita de la información contenida en un datawarehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el datawarehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el datawarehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

**No volátil:** el almacén de información de un datawarehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del datawarehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Otra característica del datawarehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo, etc.

Los metadatos serán los que permiten simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Los objetivos que deben cumplir los metadatos, según el colectivo al que va dirigido, son:

**Dar soporte al usuario final**, ayudándole a acceder al datawarehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de Business Intelligence como DSS, EIS o CMI.

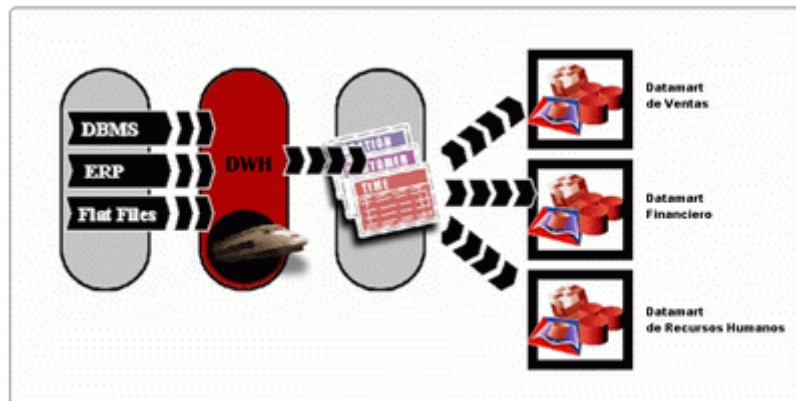
**Dar soporte a los responsables técnicos del datawarehouse en aspectos de auditoría**, gestión de la información histórica, administración del datawarehouse, elaboración de programas de extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos... etc.

Por último, destacar que para comprender íntegramente el concepto de datawarehouse, es importante entender cuál es el proceso de construcción del mismo, denominado ETL (Extracción, Transformación y Carga):

**Extracción:** obtención de información de las distintas fuentes tanto internas como externas.

**Transformación:** filtrado, limpieza, depuración, homogeneización y agrupación de la información.

**Carga:** organización y actualización de los datos y los metadatos en la base de datos.



*Figura 2.3.3.2 Ilustración de un Conjunto de Datamarts que conforman un Data Warehouse Corporativo*

### **Principales aportaciones de un datawarehouse**

Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.

Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.

Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.

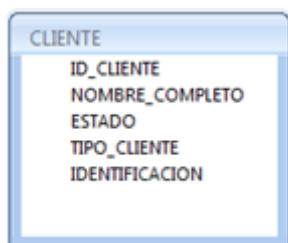
Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares.

## 2.3.4 TABLAS DE DISEÑO

A continuación una breve descripción de las tablas que serán utilizadas en el modelo de la base de datos.

### TABLA CLIENTE



*Figura 2.2.4.1*

Esta tabla representada en la Figura 2.2.4.1 muestra a todos los clientes que tiene la empresa CELLCORPO, en ella se encontrarán los datos más importantes de este, tenemos los siguientes campos:

**Id\_Cliente** el cual es un código único que distingue al cliente y que identificara a este en nuestra tabla de hecho.

**Nombre Completo:** Este campo como su nombre lo indica es donde estará el nombre del cliente.

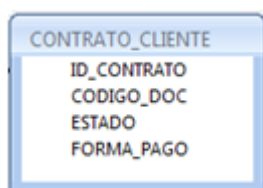
**Estado:** Este campo me sirve para conocer si un cliente está activo o inactivo en la empresa.

**Tipo Cliente:** Este campo me sirve para conocer si el cliente es persona natural (N) o jurídica (J).

**Identificación:** Este campo es para conocer la cédula o RUC del cliente.



## TABLA CONTRATO \_ CLIENTE



*Figura 2.2.4.2*

Esta tabla representada en la Figura 2.2.4.2 muestra todas las cuentas que el cliente posee dentro de la empresa CELLCORPO, ya que puede ser que un cliente tenga más de una cuenta, tenemos los siguientes campos:

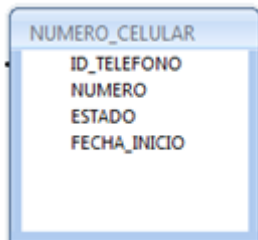
**Id\_Contrato:** Este es un código único que tendrá cada cuenta del cliente y que identificara a esta en nuestra tabla de hecho.

**Codigo\_doc:** Este campo es el número de cuenta del cliente.

**Estado:** Este campo me sirve para conocer si una cuenta se encuentra activa o inactiva en la empresa.

**Forma\_pago:** Este campo me sirve para conocer cuál es la forma de pago del cliente ya sea este contra factura, banco o tarjeta de crédito.

## TABLA NUMERO \_ CELULAR



*Figura 2.2.4.3*

Esta tabla representada en la Figura 2.2.4.3 muestra todos los números celulares que posee el cliente dentro de la empresa CELLCORPO, ya que puede ser que un cliente tenga varios números celulares, tenemos los siguientes campos:

**Id\_telefono:** Este es un código único que tendrá cada número celular y que identificara a este en nuestra tabla de hecho.

**Numero:** Este campo es el número celular del cliente.

**Estado:** Este campo me sirve para conocer si la línea del cliente se encuentra activa o inactiva en la empresa.

**Fecha\_inicio:** Este campo me sirve para conocer desde cuando el cliente inicio con su número celular en la compañía.

## TABLA DETALLE \_ PLAN

DETALLE_PLAN
ID_DETALLE_PLAN
PLAN_LINEA
DESCRIPCION
PRODUCTO
ESTADO

Figura 2.2.4.4

Esta tabla representada en la Figura 2.2.4.4 muestra todos los planes y productos que el cliente tiene dentro de la empresa CELLCORPO, para los productos son solo 2 que son PPA y AUT sin embargo en AUT existen una gama de planes que serán detallados en esta tabla, tenemos los siguientes campos :

**Id\_detalle\_plan:** Este es un código único que tendrá cada plan y que identificara a este en nuestra tabla de hecho.

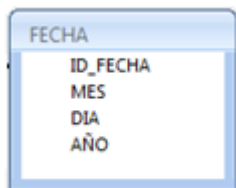
**Descripción:** Este campo me va a mostrar el nombre del plan.

**Producto:** En este campo se podrá identificar qué tipo de producto es ya sea Prepago o Autocontrol.

**Estado:** Este campo me sirve para conocer si el plan se encuentra activo o inactivo en la empresa.

**Id\_plan:** Este campo tendrá un código para los planes controlados que posee la compañía.

## TABLA FECHA



FECHA
ID_FECHA
MES
DIA
AÑO

*Figura 2.2.4.5*

Esta tabla representada en la Figura 2.2.4.5 muestra cuando el cliente empieza en la compañía, tenemos los siguientes campos:

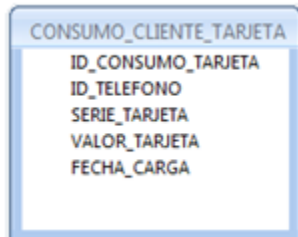
**Id\_fecha:** Este es un código único que tendrá cada fecha y que identificara a este en nuestra tabla de hecho.

**Mes:** Este campo me muestra el mes cuando el cliente empieza en la compañía.

**Día:** Este campo me muestra el día cuando el cliente empieza en la compañía.

**Año:** Este campo me muestra el año cuando el cliente empieza en la compañía.

## TABLA CONSUMO\_CLIENTE\_TARJETA



CONSUMO_CLIENTE_TARJETA
ID_CONSUMO_TARJETA
ID_TELEFONO
SERIE_TARJETA
VALOR_TARJETA
FECHA_CARGA

Figura 2.2.4.6

Esta tabla representada en la Figura 2.2.4.6 muestra todos los consumos que el cliente ha hecho vía recargas o compras de tarjetas Prepago, el consumo de estas pueden ser para clientes Prepago o clientes Autocontrol ya que una vez que este ha consumido su tarifa básica podrá hacer uso de tarjetas para poder seguir utilizando el servicio, tenemos los siguientes campos :

**Id\_consumo\_tarjeta:** Este es un código único que tendrá cada carga de tarjeta y que identificara a este en nuestra tabla de hecho.

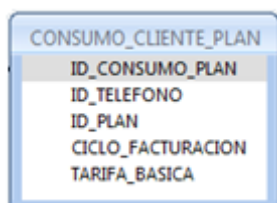
**Id\_telefono:** Este campo me va a mostrar que número celular ha hecho la recarga de tarjeta.

**Serie\_tarjeta:** En este campo se podrá identificar la serie de tarjeta que ha sido cargada por el cliente.

**Valor\_tarjeta:** Este campo me muestra el valor de la tarjeta ya que existen varias denominaciones.

**Fecha\_carga:** Este campo me muestra la fecha en que fue cargada la tarjeta.

## TABLA CONSUMO\_CLIENTE\_PLAN



CONSUMO_CLIENTE_PLAN
ID_CONSUMO_PLAN
ID_TELEFONO
ID_PLAN
CICLO_FACTURACION
TARIFA_BASICA

*Figura 2.2.4.7*

Esta tabla representada en la Figura 2.2.4.7 muestra a todos los clientes con sus planes y sus respectivas tarifas básicas, es decir, solo existirán clientes Autocontrol, tenemos los siguientes campos:

**Id\_consumo\_plan:** Este es un código único que tendrá cada consumo plan y que identificara a este en nuestra tabla de hecho.

**Id\_telefono:** Este campo me va a mostrar el número celular del cliente que se encuentra dentro de planes controlados.

**Id\_plan:** En este campo se podrá identificar el código del plan.

**Ciclo\_facturación:** Este campo me muestra en que ciclo de facturación se encuentra el cliente.

**Tarifa\_basica:** Este campo me muestra la tarifa básica del cliente según el plan que este posea.

## TABLA MAESTROS

MAESTROS
ID_CONSUMO_TARJETAS
ID_TELEFONO
PROD
ID_PLAN
DESCRIPCION
ID_FECHA
ID_DETALLE_PLAN
ID_CLIENTE
ID_CONTRATO
TOTAL_TAR
CANTIDAD_TAR
TARIFA_BASICA
TOPE_TAR_AUT
ID_CONSUMO_PLAN

Figura 2.2.4.8

Esta tabla representada en la Figura 2.2.4.8 es la tabla de hecho donde se encuentran todos los ID de todas las tablas mencionadas anteriormente y donde se muestra el resultado final de ellas, tenemos los siguientes campos:

**Id\_telefono:** Este es el código único que tendrá cada número celular.

**Id\_fecha:** Este campo es el código de fecha.

**Id\_detalle\_plan:** Este campo me muestra en que ciclo de facturación se encuentra el cliente.

**Id\_cliente:** Este campo me muestra el código del cliente.

**Id\_contrato:** Este campo me muestra el código de la cuenta del cliente.

**Id\_consumo\_tarjeta:** Este es el código del consumo por tarjeta.

**Id\_consumo\_plan:** Este es el código del consumo por plan.

**Total\_Tarifa:** Este es el valor del total de la tarifa que posee el cliente.

**Cantidad\_tarjeta:** Este es el valor del consumo por tarjetas.

**Tarifa\_basica:** Este campo me muestra la tarifa básica del cliente según el plan que este posea.

**Tope\_Tar\_Aut:** Este es el valor tope consumido de las tarjetas de plan autocontrol.

### 2.3.5 MODELO MULTIDIMENSIONAL

#### Modelo Estrella

El Modelo estrella o también conocido como Diagrama Estrella es llamado así debido a que su implementación en un ambiente relacional esta dado por varias tablas que almacenan las jerarquías dimensionales y una tabla de hecho.

El modelo estrella se presenta a continuación en el siguiente gráfico:

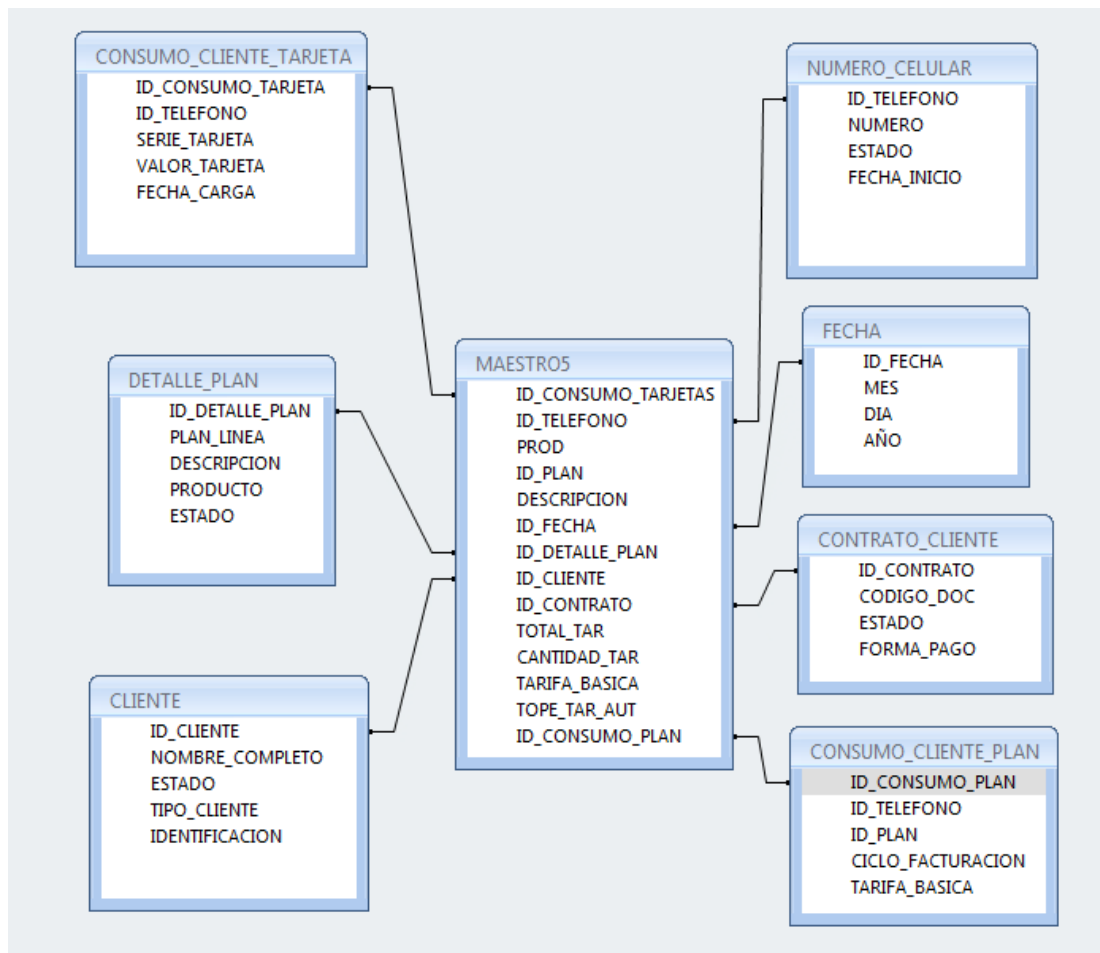


Gráfico 2.2 Diagrama Estrella del Modelo

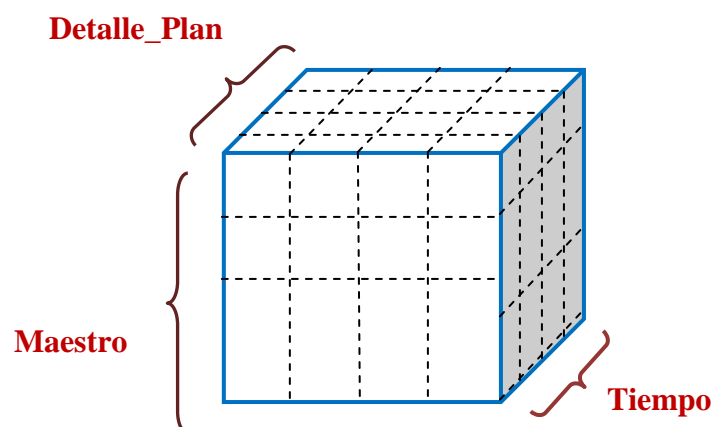


Como podemos observar en el Gráfico 2.2 las tablas de dimensiones están ligadas a la tabla de Hecho denominada Maestro, por relaciones. La integridad referencial es llevada a cabo por la creación de las llaves foráneas en la tabla Maestro, que a su vez forma parte de la llave principal de esta tabla. Es importante mencionar que las jerarquías completas son guardadas en una sola tabla dimensión.

Ya realizado el modelo estrella podemos pasar a otra definición importante en este proyecto como lo son los Cubos de Datos.

### **Cubo de Datos**

El cubo de datos es el conjunto formado por todas las tablas Dimensión y tabla de Hecho que al final dan una vista en forma de Cubo cuyas celdas están compuestas por las medidas de la tabla de Hecho. El cubo de datos es lo que hace que los reportes sean obtenidos al más corto tiempo posible y que el análisis de dichos datos pueda ser tan diverso, pues cada cara del cubo se refiere a un análisis distinto de las medidas almacenadas.



*Gráfico 2.3 Cubo de Datos*

Dada todas las tablas: clientes, consumo\_cliente\_plan, consumo\_cliente\_tarjeta, contrato\_cliente, detalle\_plan, fecha, numero\_celular junto con la tabla de hecho denominada Maestro podemos formar el cubo de datos, como se muestra en el Gráfico 2.3.

### **Observaciones al análisis efectuado**

Para el presente análisis de la información se realizaron filtros de depuración tanto para los clientes activos en la red de la compañía y adicionalmente para la carga de tarjetas (esto aplica para ambos productos PREPAGO y AUTOCONTROL).

Para clientes activos dentro de la compañía se consideraron aquellos que se encuentran en la red (activos) en los últimos tres meses para prepago y planes autocontrol. No se considero toda la base de datos de los clientes ya que contiene millones de registros y esto hace que el análisis sea más complejo.

Para carga de tarjetas se consideraron los dos últimos meses de los clientes mencionados en el párrafo anterior ya que como conocemos el producto prepago y autocontrol realiza este tipo de cargas.

Pasado los tres meses sin carga de tarjetas, se da por concluida la línea telefónica celular, es decir, se anulara dicha línea y dejara de tener servicio automáticamente.

# **CAPÍTULO III**

## **3. MODELAMIENTO DE ÁRBOLES DE DECISIÓN**

### **3.1 INTRODUCCIÓN**

En este capítulo encontramos los árboles de decisión a partir de la tabla de hecho denominada Maestro que diseñamos en el modelamiento de la minería de datos.

Aquí encontramos herramientas como SPSS AnswerTree que es un software que utiliza cuatro tipos de algoritmos para determinar por medio de árboles de decisiones la mejor variable predictora de nuestro modelo.

### **3.2 MODELAMIENTO DEL ÁRBOL**

En esta parte del capítulo III de este proyecto, determinamos y escogemos el tipo de árbol de decisión que usamos, dando a conocer las siguientes definiciones.

### 3.2.1 Árboles de decisión

Un árbol de decisión es un modelo de predicción utilizado para seleccionar una de varias alternativas posibles de decisión.

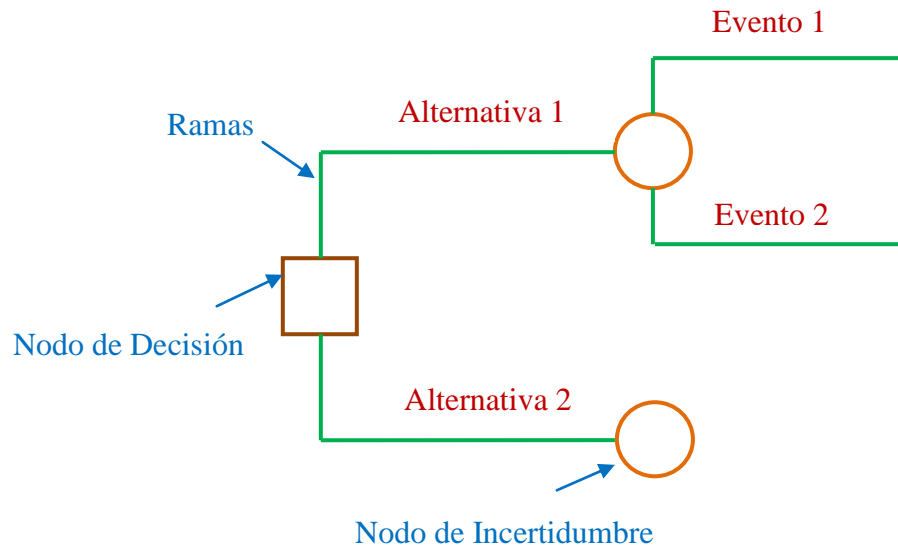
Dada una base de datos se construyen diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

El primer paso en construir un árbol de decisión es identificar las opciones que tenemos que elegir para realizar nuestros objetivos.

Los árboles de decisión poseen:

- Ramas: se representan con líneas
  
- Nodos de decisión: de ellos salen las ramas de decisión y se representan con □
  
- Nodos de incertidumbre: de ellos salen las ramas de los eventos y se representan con ○

Gráficamente un árbol de decisión puede ser representado así:

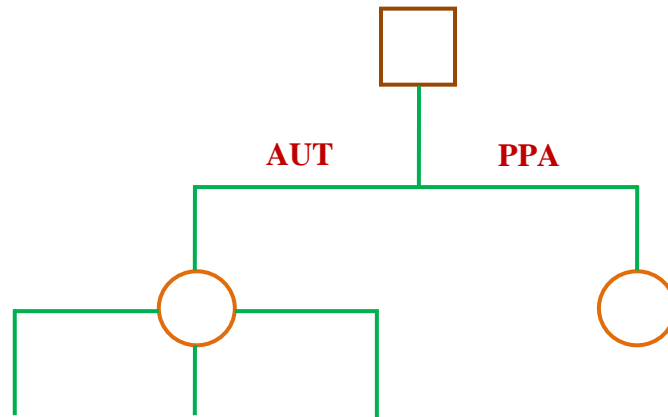


*Fig. 3.1 Representación Horizontal de un árbol de decisión*

Los árboles de decisión también suelen ser representados verticalmente, contienen los mismos componentes en su estructura y su interpretación llega al mismo resultado que en el gráfico horizontal.

En base a la teoría ya mencionada anteriormente, observamos que nuestras alternativas son los dos tipos de planes a decidir plan Pre-pago (PPA) y plan Autocontrol (AUT), que a su vez son dos ramificaciones del NODO DE DECISIÓN, por tanto tendremos dos NODOS DE INCERTIDUMBRE.

Gráficamente parte de nuestro árbol de decisión es:



*Fig. 3.2 Representación Vertical del Árbol de Decisión*

Para el análisis del árbol de decisión nos valemos de un software estadístico, denominado SPSS AnswerTree Versión 3.

**SPSS AnswerTree** es una herramienta para la creación de sistemas de clasificación en forma de árbol de decisión, lo cual nos ofrece una fácil interpretación de los resultados. Permite crear perfiles de clientes y segmentar bases de datos, sus aplicaciones son muy variadas pero podemos destacar la valoración de créditos, mejora de los resultados de campañas de marketing directo, investigación médica, detección de operaciones fraudulentas, análisis de venta cruzada, identificación de clientes potenciales, etc.

A diferencia de otros métodos de clasificación como el Análisis Clúster, **AnswerTree** nos permite realizar tipologías de clientes en función de un variable criterio, así como realizar pronósticos con

probabilidades conocidas, siendo por esto, considerada como una herramienta con gran poder predictivo.

**AnswerTree** lee datos en distintos formatos a través de conexiones ODBC estándares y puede ser ejecutado como un programa independiente o integrado dentro de SPSS Base.

Su sencillez le permitirá comenzar a crear modelos rápidamente gracias a su asistente de árbol, con el cual, en 4 sencillos pasos, definirá por completo su análisis. Lo único que debemos decidir será:

- ¿Qué algoritmo seleccionar?
  
- ¿Cuál será la variable dependiente también llamada target u objetivo? Por ejemplo, si responde a cierto mailing o no, si devuelve un crédito o no, etc.
  
- Seleccionar los predictores o variables independientes. Por ejemplo, edad, sexo, salario, categoría laboral, etc. que nos ayudarán a crear los perfiles.

Una vez decidido esto, **AnswerTree** generará el árbol para nosotros; este modelo puede ser aplicado a nuestra base de datos con el fin de generar listados de clientes susceptibles de devolver un crédito, o de adquirir cierto producto, etc.

Ya escogidos los datos de la tabla de hecho, debemos escoger el tipo de árbol de decisión a utilizar para la poda y tener como resultado la predicción.

AnswerTrees incluye cuatro algoritmos de árboles establecidos:

**CHAID:** un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Un método que utiliza las estadísticas Ji-Cuadrado o la F para predictores de selección. Cada división puede tener varios nodos.

**CHAID exhaustivo:** una modificación de CHAID que examina todas las divisiones posibles para cada predictor. Toma más tiempo para calcular.

**Árboles de clasificación y regresión (C&RT):** un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos.

**QUEST:** un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz.

Con cuatro algoritmos, tiene la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a sus datos.

Con los datos de la tabla de Hecho, Maestro, la cual contiene doscientos veinte y tres mil doscientos cincuenta y seis registros, y con ayuda del software antes mencionado ingresamos dichos registros para hacer el árbol de decisión, podemos por medio del árbol tipo CHAID para realizar la siguiente predicción.



El uso de esta técnica supone distinguir, por un lado, una variable cuya distribución se desea explicar y, por el otro, un conjunto de variables nominales u ordinales con el status de independientes. Estas últimas reciben el nombre de pronosticadoras y tienen el cometido de formar grupos que sean muy distintos entre sí en la variable dependiente o criterio. En este caso las variables predictoras son cantidad y Total Tarifa, es decir las variables independientes para ambos planes. Aunque también encontramos dos variables más que son tarifa básica y tope tarifa autocontrol, pero dichas variables solo son utilizadas en plan autocontrol, aunque con la restricción de que durante tres meses debes al menos ingresar una tarjeta en tu plan pre-pago para que tu línea celular no sea anulada, una de dichas variables puede ser utilizada como lo veremos más adelante.

Para efectos del análisis de segmentación procedimos a definir como variable dependiente (criterio), el tipo de plan que el cliente posee, denominada Producto (PROD). Como variables predictoras se tomaron en cuenta: la cantidad y la total tarifa. Estas variables se midieron en un nivel cuantitativo. Cada una de ellas se ha dividido en dos para ser consistente con los requerimientos del algoritmo de segmentación utilizado, Chaid (Chi-Squared Automatic Interaction Detection), el cual utiliza el estadístico  $X^2$  para seleccionar los mejores predictores y permite que la muestra se segmente en más de dos categorías, por lo que cada división puede dar lugar a múltiple nodos (Escobar, 1998; SPSS AnswerTree 3.1, 2002).

Ya escogida y definida las variables procedemos a definir los intervalos en el que se analizará y resumirá el árbol de decisión para hallar la mejor variable predictora.

En la Tabla I se presenta la dicotomización de las variables predictoras o independientes del primer árbol a analizar, tomando en cuenta las variables predictoras utilizables en ambos planes, tanto autocontrol como pre-pagos.

**TABLA I**  
ANÁLISIS DE SEGMENTACIÓN. VARIABLES PREDICTORAS

<b>VARIABLES PREDICTORAS</b>	<b>VALORES</b>
Cantidad	Menores o igual a 14 tarjetas ingresadas. Mayores a 14 tarjetas ingresadas
Total Tarifa	Menor o igual a 98 dólares consumidos. Mayores a 98 dólares consumidos.

El uso de esta técnica de segmentación supone distinguir, por un lado, una variable cuya distribución se desea explicar y, por el otro, un conjunto de variables nominales u ordinales con el status de independientes. Estas últimas reciben el nombre de pronosticadoras y

tienen el cometido de formar grupos que sean muy distintos entre sí en la variable dependiente o criterio. En este caso nuestras variables predictoras son Cantidad, Total Tarifa, es decir las variables independientes, estas variables son cuantitativas.

**TABLA II**  
*PREDICTORES DE LA VARIABLE: PRODUCTO*

<b>Predictores</b>	<b>Chi-Cuadrado (<math>X^2</math>)</b>	<b>Diferencia Significativa</b>
Cantidad	273,48	0,00
Total Tarifa	22,73	0,00

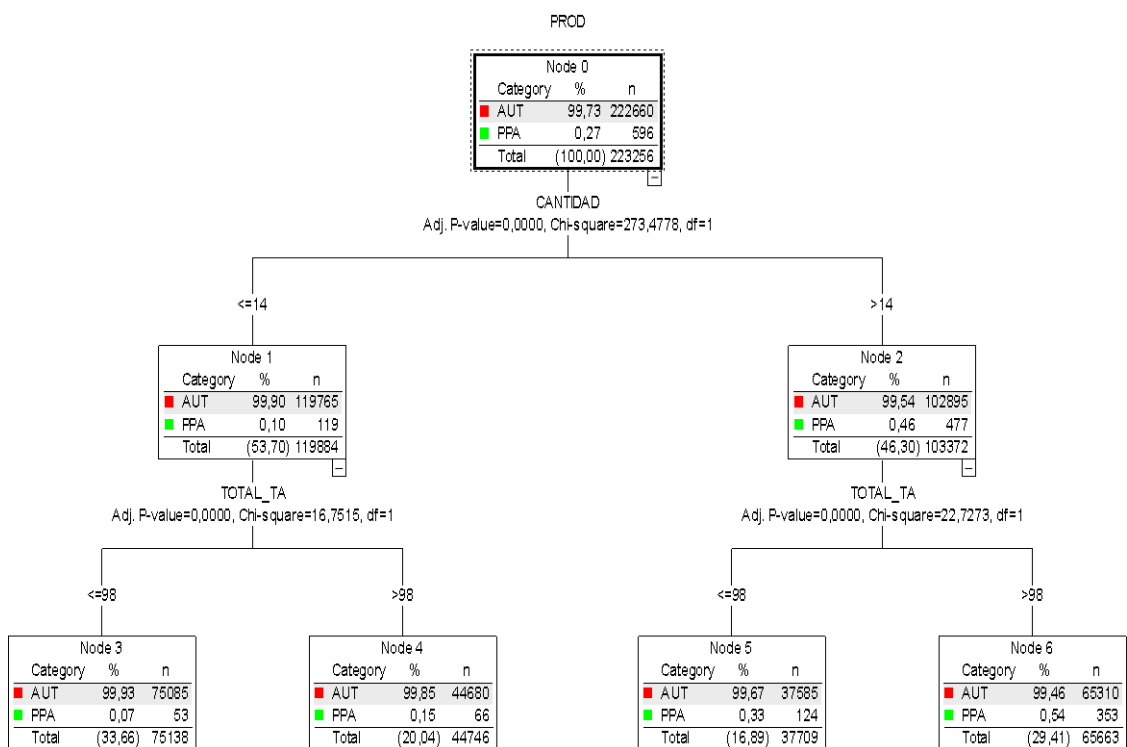
En la Tabla II arriba indicada se muestran por orden de importancia los mejores predictores del variable producto. Se incluyen Cantidad y Total Tarifa. Estas son las dos primeras variables que mejor explican la variable criterio, en virtud de que tienen los valores Chi-Cuadrado más altos.

El árbol de segmentación general que se muestra en el Gráfico 3.1 se divide en dos niveles a partir del nodo parental (nodo 0). Cada nivel está representado por una variable predictora colocada en orden de prioridad. De hecho, la cantidad resulta ser el mejor predictor, por tener el  $X^2$  más

alto (273,48). De igual forma, es la variable en la que se observa mayor diferencia estadísticamente significativa entre los grupos de categorías que la integran, como es el caso de las menores y mayores de 14 tarjetas ingresadas durante los tres últimos meses.

### GRÁFICO 3.1

#### ÁRBOL DE DECISION. PREDICTOR: Cantidad



Fuente: AnswerTree SPSS 3.0

Elaborado: Lorena Calvas y Ángela Vásconez

En el Gráfico 3.1 se muestra el árbol de decisión siendo para ambos planes la cantidad el mejor predictor, las variables cantidad y total tarifa fueron definidas con dos intervalos, los cuales podemos observar en la Tabla 1 los valores de dichas variables.

El nodo raíz muestra el desglose de los casos de toda la muestra (223.256 clientes).

En este conjunto de datos se observa que los clientes poseen plan autocontrol en su mayoría (99,73%) y el resto plan prepago (0,27%).

Cabe destacar que la mayoría de los clientes que ingresan la cantidad menor o igual a catorce tarjetas durante los tres meses poseen plan autocontrol (99,90%), así mismo los clientes que ingresan mayores de catorce tarjetas durante los tres meses poseen plan autocontrol (99,54%).

El estudio se realiza durante tres meses, es por ello que cabe recalcar que pasado este período de tiempo los clientes que poseen planes pre-pagos si no han ingresado minutos de tiempo aire en su línea celular, dicha línea será anulada totalmente. La caducidad de las tarjetas se muestra a continuación:

**Tiempos de vigencia:**

**\$3:** 7 días - **\$6:** 30 días - **\$10:** 30 días - **\$20:** 45 días - **\$30:** 45 días

Es decir que para mantener activa su línea celular debe haber realizado durante los últimos tres meses al menos una recarga a su plan pre-pago, de lo contrario perderá su número o línea celular.

### **3.3 SOLUCIÓN DEL PROBLEMA**

Con la segmentación ya realizada y guiándonos por las variables predictoras que el árbol de decisión nos dio llegamos a sesenta y cinco mil seiscientos sesenta y tres a los cuales podemos ayudar con la mejora del plan que posee.

# CONCLUSIONES

1. Realizando la segmentación respectiva llegamos a la conclusión de que de los doscientos mil registros aproximados sesenta y cinco mil seiscientos sesenta y tres son los clientes a los cuales debemos “atacar”, proponiendo un nuevo plan que le darán mayores beneficios que el plan que tiene actualmente.
2. De los sesenta y cinco mil seiscientos sesenta y tres clientes, más del 90% de los clientes poseen un plan autocontrol según análisis de estudio.
3. De los sesenta y cinco mil seiscientos sesenta y tres clientes, menos del 10% de los clientes poseen un plan prepago según análisis de estudio.
4. El 30% de los clientes son aquellas personas que deben cambiar su plan por una mejor opción, debido a su hábito de consumo y así ahorrarían dinero.
5. La variable Cantidad de tarjetas es la que nos ayuda a predecir por medio de árboles de decisión los clientes que deben cambiar de plan.

6. Los clientes que son resultado de nuestro estudio ingresan más de catorce tarjetas en el trimestre.
  
7. Los clientes que son resultado de nuestro estudio consumen más de noventa y ocho dólares en el trimestre.



## **RECOMENDACIONES**

1. Los clientes que son resultado del estudio deben cambiar a un nuevo plan de telefonía celular.
2. Se debería tomar en cuenta la base de datos la tarifa promedio que utilizan todos los clientes en ambos planes, así sería más rápido obtener a los clientes que mas ingresan tarjetas durante los tres meses tanto en plan autocontrol como en planes prepagos.
3. Sería de mucho beneficio que la empresa realice una prueba piloto con los clientes que han sido resultado del estudio, para que ellos se den cuenta del gran beneficio que ellos lograrían cambiando su plan actual. La prueba piloto se basaría en hablar con dichos clientes y con la aprobación de ellos cambiar al plan que se le recomienda por al menos tres meses y así ellos podrían comparar con los tres meses anteriores.

# BIBLIOGRAFÍA

1. SPSS (2002). Answer Tree 2.0. User's Guide. Chicago: SPSS. Statistics Package for the Social Sciences (2002). Answer Tree 2.0. User's Guide. Chicago: SPSS.
2. Statistics Package for the Social Sciences (2002). Answer Tree 2.0. User's Guide. Chicago: SPSS.
3. 2008, [http://www.google.com.ec/imgres?imgurl=http://www.ingelsi.com.ec/spss/marketing-3.gif&imgrefurl=http://www.ingelsi.com.ec/spss/marketing-pack.htm&usq=\\_mK11nN-M6RiHYsXVLRdj-qmEC3U=&h=375&w=500&sz=18&hl=es&start=6&um=1&itbs=1&tbnid=x3XajYQFb\\_j3BM:&tbnh=98&tbnw=130&prev=/images%3Fq%3Dinterpretacion%2Bde%2Blos%2Bresultados%2Bde%2Bbarboles%2Bde%2Bdecision%2Bpor%2Banswertree%26um%3D1%26hl%3Des%26sa%3DN%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.ingelsi.com.ec/spss/marketing-3.gif&imgrefurl=http://www.ingelsi.com.ec/spss/marketing-pack.htm&usq=_mK11nN-M6RiHYsXVLRdj-qmEC3U=&h=375&w=500&sz=18&hl=es&start=6&um=1&itbs=1&tbnid=x3XajYQFb_j3BM:&tbnh=98&tbnw=130&prev=/images%3Fq%3Dinterpretacion%2Bde%2Blos%2Bresultados%2Bde%2Bbarboles%2Bde%2Bdecision%2Bpor%2Banswertree%26um%3D1%26hl%3Des%26sa%3DN%26tbs%3Disch:1)
4. 2008, [http://www.telematica.com.pe/Product/spss\\_AnsTree.htm](http://www.telematica.com.pe/Product/spss_AnsTree.htm)
5. 2009, [http://es.wikipedia.org/wiki/Minería\\_de\\_datos](http://es.wikipedia.org/wiki/Minería_de_datos)
6. 2009, [http://www.answermath.com/mineria\\_de\\_datos.htm](http://www.answermath.com/mineria_de_datos.htm)
7. 2009, [http://www.aprendedynamics.com/conceptos\\_olap.html](http://www.aprendedynamics.com/conceptos_olap.html)
8. 2009, <http://www.slideshare.net/dannoblack/datawarehouse-y-datamining-parte-i>

9. 2009, [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cube.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cube.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=__2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1)
10. [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cube.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cube.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=__2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1)
11. 2009, [http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart\\_cube.jpg&imgrefurl=http://www.sinnexus.com/business\\_intelligence/datamart.aspx&usq=\\_\\_2MHYfKY5SO\\_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ\\_ESEC377%26tbs%3Disch:1](http://www.google.com.ec/imgres?imgurl=http://www.sinnexus.com/images/datamart_cube.jpg&imgrefurl=http://www.sinnexus.com/business_intelligence/datamart.aspx&usq=__2MHYfKY5SO_aAA9sCJFus8FuhpQ=&h=167&w=199&sz=6&hl=es&start=3&um=1&itbs=1&tbnid=ua2NqymvrxmzsM:&tbnh=87&tbnw=104&prev=/images%3Fq%3Ddatamart%26um%3D1%26hl%3Des%26sa%3DN%26rlz%3D1G1GGLQ_ESEC377%26tbs%3Disch:1)