

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**GENERACIÓN DE UN MODELO DE PROPENSIÓN DE COMPRA EN LA
INDUSTRIA DE LAS TELECOMUNICACIONES (CASO ENTEL SUSCRIPCIÓN
PERSONAS)**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

MAXIMILIANO ANDRÉS ALVARADO NARANJO

PROFESOR GUÍA:
CRISTIAN MATURANA

PROFESOR CO-GUÍA:
MÁXIMO BOSCH

PROFESOR INTEGRANTE:
M. ÁNGELES DÍAZ

SANTIAGO DE CHILE
ABRIL 2012

RESUMEN DE LA MEMORIA
PARA OPTAR AL TITULO DE
INGENIERO CIVIL INDUSTRIAL
POR: MAXIMILIANO ALVARADO NARANJO
FECHA: 25/04/2012
PROF. GUIA: SR. CRISTIAN MATURANA

GENERACIÓN DE UN MODELO DE PROPENSIÓN DE COMPRA EN LA INDUSTRIA DE LAS TELECOMUNICACIONES (CASO ENTEL SUSCRIPCIÓN PERSONAS)

El contexto actual de la industria de las telecomunicaciones ha estado marcado por un alto grado de competencia, acelerados cambios tecnológicos, rebajas tarifarias, portabilidad numérica, solidez y una gran rentabilidad. Por estas razones, Entel, se ha propuesto masificar la experiencia multimedia y aumentar el ingreso promedio por usuario.

El objetivo de esta memoria es generar un modelo de propensión de compra para crear ofertas focalizadas a usuarios que no hayan realizado up-selling a Planes Multimedia. Para ello, se seleccionó un conjunto de datos objetivo, los cuales fueron procesados mediante herramientas de Minería de Datos (Data Mining), donde se evaluaron y se seleccionaron los modelos predictivos con mayor ganancia de información. El paso siguiente fue identificar patrones y caracterizar a los clientes más propensos a realizar el up-selling. Por último, este trabajo contempla una evaluación económica que mide el impacto que podría llegar a tener la implementación de las ofertas focalizadas.

La metodología utilizada se basa en herramientas de Minería de Datos, específicamente, se utilizó la metodología KDD (Knowledge Discovery for Data Base). Con esta metodología, se buscó descubrir patrones ocultos dentro de cantidades masivas de datos.

En una etapa previa a la modelación se analizaron los datos descriptivamente, donde las variables más relevantes fueron la edad, el género y sector económico.

Luego se aplicaron los modelos predictivos, Árbol de Decisión, Regresión Logística y Red Neuronal. La mejor predicción se obtuvo con Árbol de Decisión que entrego un 83,39% de exactitud en la clasificación. Dentro de las variables más relevantes para predecir el up-selling se encuentra el consumo de descarga de internet móvil, índices financieros dentro de la compañía, tráfico de mensajes, edad, nivel de facturación, tráfico de minutos emitidos y contratación de bolsas.

Por último se realizó una etapa de evaluación económica, a partir de los ingresos futuros que generarían los clientes propensos a realizar el up-selling a Planes Multimedia. Considerando, solamente el 10% de los datos, los beneficios esperados son \$30.652.312.

AGRADECIMIENTOS

Quisiera expresar mis agradecimientos a la Compañía Entel S.A., en especial a M. Ángeles Díaz, Brenda Zarate y Santiago Polanco, por haberme dado la oportunidad de concluir esta etapa.

Además, quisiera agradecer a los profesores Cristian Maturana y Máximo Bosch, por haber cooperado durante todo el periodo en que se realizó este trabajo con mucha disponibilidad y comprensión.

Dedicada a mi padre Dagoberto
Alvarado que se encuentra
observando, desde algún lugar,
el fin de la etapa que siempre
quiso ver y a mi abuela Cecilia
Miguez.

De manera muy especial para
mi madre Zaida Naranjo, para
mi hermano Dagoberto Alvarado
Naranjo y para Magdalena
Ceballos.

Índice de Contenido

1. ANTECEDENTES GENERALES	7
2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN	10
3. OBJETIVOS	13
3.1 OBJETIVO GENERAL.....	13
3.2 OBJETIVOS ESPECÍFICOS.....	13
4. MARCO CONCEPTUAL.....	13
4.1 DATA MINING (Minería de Datos).....	13
4.1.1 MODELOS PREDICTIVOS.....	15
5. METODOLOGÍA.....	25
6. ALCANCES	28
7. RESULTADOS ESPERADOS	29
8. DESARROLLO DE LA METODOLOGÍA.....	29
8.1 RECOPIACIÓN DE LAS BASES DE DATOS	29
8.2 COMPRENSIÓN DE LOS DATOS	32
8.3 PREPARACIÓN DE LOS DATOS	33
8.4 EXPLORACIÓN DE LOS DATOS.....	38
8.4.1 TIPOS DE COMPORTAMIENTOS EN LOS PLANES	40
8.4.2 TIPOS DE COMPORTAMIENTOS EN LOS PLANES DE UP-SELLERS....	43
8.4.3 ANÁLISIS DEMOGRÁFICO Y GEOGRÁFICO DE LOS UP-SELLERS.....	45
8.4.4 ANÁLISIS DE FACTURACIÓN DE UP-SELLERS	49
8.4.5 ANÁLISIS DEL TRÁFICO DE MINUTOS DE LOS UP-SELLERS	51
8.4.6 ANÁLISIS DEL TRÁFICO DE MENSAJES DE LOS UP-SELLERS	52
8.4.7 ANÁLISIS DE EQUIPOS CELULAR DE LOS UP-SELLERS	54
8.4.8 RESULTADOS EN LA COMPRENSIÓN DE LOS DATOS	56
8.5 MINERÍA DE DATOS	57
8.5.1 APLICACIÓN DE ÁRBOL DE DECISIÓN A CLIENTES SUS	57
8.5.2 APLICACIÓN DE ÁRBOL DE DECISIÓN A CLIENTES CC	66
8.5.3 APLICACIÓN DE RED NEURONAL A CLIENTES SUS	73
8.5.4 APLICACIÓN DE RED NEURONAL A CLIENTES CC	77
8.5.5 APLICACIÓN DE REGRESIÓN LOGÍSTICA A CLIENTES SUS	79

8.5.6	APLICACIÓN DE REGRESIÓN LOGÍSTICA A CLIENTES CC	82
8.6	EVALUACIÓN ECONÓMICA.....	84
8.7	PRESENTACIÓN DEL CONOCIMIENTO Y CONCLUSIONES	88
9.	BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN	90
10.	ANEXOS	92

1. ANTECEDENTES GENERALES

La industria de las telecomunicaciones tiene por objetivo comunicar, es decir, transmitir voz, datos e imágenes con una calidad similar a la de dos personas que se encuentran a un metro de distancia. En los últimos años, la demanda por la red de telecomunicaciones ha tenido un aumento, en especial por los servicios de Internet. En este sentido, la actual preocupación es controlar la red (conjunto organizado de recursos que son compartidos por todos los usuarios) para permitir un intercambio de información constante y continuo. Dentro de los servicios ofrecidos se encuentra la telefonía fija y móvil, la banda ancha fija y móvil, y la televisión (por cable o satélite).

En la última década, la industria de las telecomunicaciones se ha transformado en uno de los negocios más dinámicos del mundo, debido a los acelerados cambios tecnológicos y al alto grado de competencia.

La crisis financiera, del 2007-2008, sacudió bruscamente a gran parte de las economías e industrias mundiales. Esto se tradujo en que el comercio internacional se desmoronara, el capital dejara de circular y el crecimiento se enfriara. Pero lo anterior, no se vio reflejado en la industria nacional de las telecomunicaciones, ya que el año 2009 tuvo un crecimiento del 6,9% con respecto a su periodo anterior¹. La principal razón de obtener aquellos resultados se refleja en que la telefonía celular e Internet se han convertido en una necesidad para la mayoría de los chilenos.

En particular, el año 2010, las utilidades de Entel S.A llegaron a los MM\$172.971. Cifra que es superior en un 22% a las utilidades consolidadas del año anterior. Los negocios de redes móviles aportaron con el 77% de los ingresos (Anexo 1). Además, el mismo año, la compañía amplió su base de clientes a más de 7,5 millones de usuarios² (Anexo 2).

El desafío de la industria, en las últimas dos décadas, fue elevar la penetración de mercado de la telefonía móvil, que en el año 2010 llegó al 119% en nuestro país (Anexo 3). Es así, como Chile posee uno de los índices de penetración de telefonía móvil más altos en Latinoamérica, índice mayor que la del resto de los países de la región que poseen un PIB per cápita superior al de nuestro país³ (Anexo 4). En base a lo anterior, el desafío actual por incrementar los ingresos de las compañías de telecomunicaciones se traduce en la generación de nuevos productos y el desarrollo de la propia base de clientes. En definitiva, la industria de las telecomunicaciones se ha planteado el nuevo desafío de masificar la experiencia multimedia.

¹ Fuente: Informe Anual SUBTEL, 2009

² Fuente: Memoria Anual Entel S.A., 2010

³ Fuente: Memoria Anual Entel S.A., 2010

Dentro de las principales tendencias del mercado de la telefonía móvil se manifiestan los dispositivos móviles con acceso a Internet (Smartphone y Tablet), el auge del video sobre redes móviles, el desarrollo de aplicaciones y las nuevas formas de acceso a contenidos de red. Estas tendencias están incrementando considerablemente la comunicación y productividad de las personas y las empresas, extendiéndose a redes sociales en las cuales se puede compartir una gran cantidad de información.

Además, sucesos como rebajas tarifarias, aumento en el servicio de Internet Móvil, la portabilidad numérica y el ingreso de nuevos actores han hecho más intensa la competencia dentro de la industria de la telefonía móvil.

Por lo tanto, estas condiciones de mercado y las nuevas tendencias, hacen bastante atractivo el negocio y obligan a que las empresas tomen decisiones para obtener ventajas competitivas y acceder a rendimientos sobre el promedio. Sobre todo, porque la Telefonía Móvil y el Internet Móvil se han convertido en el segmento más importante de la industria de las Telecomunicaciones, y más aún, son unos de los principales protagonistas del desarrollo y crecimiento de la economía chilena⁴.

Durante muchos años Entel PCS fue la filial móvil de Entel S.A., actualmente existe un área de Telefonía Móvil dentro de la compañía tras haber sufrido una transición ligada a una integración de empresas. Actualmente el área de Telefonía Móvil está dividida en tres segmentos de clientes: Mercado Personas, Mercado Empresas y Mercado Corporaciones.

Entel Telefonía Móvil posee una base de clientes cercana a los 7,5 MM. El número de abonados para el Segmento Pre-Pago es superior a 5 millones y en el caso del Segmento Pos-Pago (Suscripción y Cuenta Controlada) es cercano a 2,5 millones⁵ (Anexo 5).

El segmento target para la realización del proyecto, se define como los clientes que pertenecen al Mercado Personas y además que tengan contratado un plan Suscripción o un plan Cuenta Controlada (aprox. 1,3 MM⁶). La decisión se basa en que el área donde se realiza el proyecto, corresponde al Área de Ventas del Mercado Personas. La decisión de elegir a este segmento de clientes se basa en la contratación de servicios recurrentes, por lo que los esfuerzos sobre estos segmentos tienen mayor retorno y la variedad de productos a ofrecer es mucho mayor que la de clientes Pre Pago.

⁴ Fuente: Memoria Anual Entel S.A., 2010

⁵ Fuente: Estadísticas Telefonía Móvil SUBTEL, 2010

⁶ Fuente: Memoria Anual Entel S.A., 2010

Otro de los objetivos de la compañía es aumentar los ingresos promedios por usuarios (ARPU). Para esto, se destaca la iniciativa de incrementar el valor de los clientes dentro de la compañía, ofreciendo productos complementarios (cross-selling) y/o ofreciendo productos más avanzados (up-selling).

Dentro de los principales servicios que ofrece la compañía se destaca el Servicio de Voz (planes y bolsas) en modalidad Pre-Pago y Pos-Pago, Servicio de Banda Ancha Móvil (planes y bolsas) en modalidad Pre-Pago y Pos-Pago, Servicio de Internet Móvil en modalidad Pre-Pago y Pos-Pago, Servicios de Valor Agregado (soluciones de correo BlackBerry, mensajería, roaming, etc.), Servicios de Banda Ancha Fija, Televisión Digital Home y Equipos Terminales con acceso a Internet.

La investigación se centró en los Planes de Voz e Internet Móvil. La elección de los servicios nombrados se debe a que Entel se ha propuesto masificar la experiencia multimedia y aumentar los ingresos promedio por usuario.

Para poder aumentar el valor de los clientes, es necesario utilizar la información almacenada en las bases de datos. En particular, herramientas como el cross-selling y el up-selling son relevantes en la predicción de compras futuras. De esta manera, fue necesario encontrar información que explique la diferencia entre los clientes que han contratado un Plan Multimedia y los que aún no lo han hecho.

La novedad de este trabajo, radica en que además de analizar la propensión de compra a un servicio, se analizará el up-grade de compra de paquetes de productos y de las distintas modalidades que puede tomar cada uno de ellos (planes).

En síntesis, el presente trabajo tiene como objetivo generar modelos de predicción de compra para aumentar los ingresos de la compañía y satisfacer las necesidades de los clientes. Es así, como se generarán ofertas focalizadas a clientes de baja vinculación (clientes que consumen pocos productos y no se han desarrollado con los servicios de la compañía) para realizar up-Selling.

2. DESCRIPCIÓN DEL PROYECTO Y JUSTIFICACIÓN

La empresa de telecomunicaciones en la cual se desarrolló la investigación es Entel S.A., específicamente en el Área de Ventas, la cual está encargada de planificar, coordinar y controlar el cumplimiento de estrategias de comercialización, desarrollando una gestión eficiente y rentable, para asegurar así el desarrollo y crecimiento de la compañía en el largo plazo, y lograr que Entel sea el preferido en atención y ventas.

Teniendo en cuenta el rol del Área de Ventas, las tendencias del mercado y la alta competencia nace la inquietud de aumentar el valor de los clientes que posee la empresa y satisfacer sus necesidades, con el objetivo de incrementar los ingresos promedios por usuario y las utilidades de la compañía. La manera de lograrlo es creando ofertas focalizadas a través de herramientas de CRM (cross-selling y up-selling).

En consecuencia, se buscó generar un modelo de propensión de compra para aumentar los ingresos promedios por usuarios del Mercado Persona Pos-Pago de Entel Telefonía Móvil.

Fue fundamental definir un conjunto de datos objetivo en base a datos transaccionales, geográficos, demográficos, comportamiento crediticio, etc. La metodología utilizada en el desarrollo de la investigación se basa en herramientas de Minería de Datos (Data Mining), específicamente en la metodología KDD (Knowledge Discovery for Data Base). Esta metodología permite descubrir patrones ocultos en grandes cantidades de datos. En base a lo anterior, se recopilaron las bases de datos, de manera de comprender, preparar, seleccionar y transformar los datos para la explotación de minería de datos.

Luego se aplicaron las técnicas de Data Mining, como Árboles de Decisiones, Regresiones Logísticas y Redes Neuronales. Al momento de capturar que técnica es la más adecuada o la que entrega mejores resultados, se procedió a implementar la técnica para encontrar los patrones ocultos dentro de las características de los consumidores y obtener las reglas que detectan a los clientes más propensos realizar un up-selling.

Los Planes Multimedia comprenden tres servicios fundamentales, el servicio de VOZ, el servicio de DATOS (internet móvil) y el servicio de MENSAJERÍA. Estos planes son los que actualmente están dejando mayor valor dentro de la compañía y son los planes hacia donde se está moviendo el negocio.

Por último, para poder cuantificar el impacto de este proyecto, fue necesario evaluar económicamente los beneficios de aplicar estas ofertas, tomando como

hipótesis que al aplicar las ofertas, los ingresos promedios por usuario aumentarán en relación a la eficiencia y eficacia de las ofertas.

La participación de mercado en Telefonía Móvil, el año 2010, deja a Entel en un segundo lugar dentro de la industria. Movistar registra un 40,95%, de los abonados, siguen en orden de importancia, Entel con un 36,42%, Claro con 22,49%, Nextel con 0,14% e Interexport con un 0,003%. La participación de mercado del negocio Internet Móvil, se comporta de manera similar: Movistar registra un 42%, de los abonados, siguen en orden de importancia, Entel con un 31% y Claro con un 27%⁷ (Anexo 6). Por lo tanto, la entrada de nuevos operadores y las características del mercado hacen que la competencia siga en crecimiento y que los nuevos operadores luchen por obtener mayor participación de mercado.

El servicio VOZ (servicio básico que ofrece la compañía) se transformó en un mercado altamente competitivo que ha ido presentando menores tasas de crecimiento debido al gran índice de penetración de telefonía móvil. Por lo tanto es necesario desarrollar la base de clientes y potenciar nuevas fuentes de crecimiento como los Servicios de Valor Agregado.

La Banda Ancha Móvil (BAM) ha sido uno de los productos de mayor crecimiento en los últimos años. El año 2008 la base de clientes BAM era de solo 126.442, y el año 2010 la base de clientes BAM llegó a ser 550.879⁸ (Anexo 7). Este comportamiento valida la necesidad de los clientes por estar en constante conexión con la Web y las redes sociales.

Variables como el crecimiento de la base de abonados de la compañía, la continúa innovación, el desarrollo en Servicios de Valor Agregado y un fuerte crecimiento en el segmento post-pago han influido en que el ARPU se haya mantenido relativamente constante a través de los años. El año 2008, se obtuvo un ARPU de \$10.088 y el año 2010 se obtuvo un ARPU de \$9.460⁹ (Anexo 8). En este sentido la industria ha utilizado economías de escalas donde no se es tan intenso en inversión, la competitividad ha aumentado y se han rebajado costos implicando en precios inferiores y por lo tanto los ingresos también.

Es común que en las compañías de telefonía móvil entreguen un exceso de ofertas generalizadas a través de medios de comunicación poco eficientes, obteniendo así una baja captación. Para esto es posible plantear como hipótesis que la estrategia para ofrecer los servicios no es la correcta o que los planes se ofrecen de una forma

⁷ Fuente: Estadísticas Telefonía Móvil e Internet SUBTEL, 2010

⁸ Fuente: Memoria Anual Entel S.A., 2010

⁹ Fuente: Memoria Anual Entel S.A., 2010

inadecuada al público objetivo. Y es aquí donde el trabajo aporta con un mayor grado de exactitud al momento de ofrecer servicios a los clientes indicados.

Entel S.A. ha sido elegido como el número uno en el Índice Nacional de Satisfacción de Consumidores ocho años seguidos. Pero una de las críticas que se le hace a la industria de las telecomunicaciones, es que lejos de mejorar como sector, ha bajado sus evaluaciones. Y una de las variables que ha hecho que la industria no se derrumbe es la creciente tendencia a la “paquetización” (ofertas que incluyen varios productos) y a la complejidad cada vez mayor de los distintos servicios contratados¹⁰. Punto clave de esta memoria, puesto que lo que se desea ofrecer, de manera correcta a través de ofertas focalizadas a grupos de clientes, es la “paquetización” de los servicios de VOZ y DATOS.

En conclusión, se desea afinar la selección de ofertas para poder satisfacer a los clientes de la manera adecuada y seguir desarrollando una cercanía ligada a la calidad del servicio. El punto de partida para aquello es conocer el comportamiento de los clientes, sus preferencias y sus necesidades. Este trabajo realizó un estudio del comportamiento histórico de los clientes para poder identificar y predecir su comportamiento futuro.

Si bien, hoy en día, existe una gran gama de planes, bolsas y productos donde los clientes se pueden autoseleccionar según sus preferencias, no se posee un estudio exhaustivo de los clientes para determinar patrones ocultos en su comportamiento. De esta forma, al poder saber cómo se comportan y qué es lo que buscan, se podría enfrentar nuevas oportunidades de negocio en relación a ofertas individuales o de grupos de clientes que posean características similares, obteniendo como resultado final la entrega del producto y servicio correcto al cliente correcto en el tiempo correcto, aumentando los márgenes de ganancias y extrayendo el excedente del consumidor.

¹⁰ Fuente: Ranking Calidad de Servicio 2011, Pro Calidad

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Generar un modelo de propensión de compra que genere ofertas focalizadas a clientes de baja vinculación a la compañía a través del up-selling a Planes Multimedia.

3.2 OBJETIVOS ESPECÍFICOS

- Identificar perfiles de clientes de manera descriptiva, propensos a realizar up-selling a Planes Multimedia.
- Generar y evaluar distintas técnicas de Data Mining ha implementar para determinar la propensión de compra.
- Identificar patrones y caracterizar a los clientes más propensos a comprar Planes Multimedia.
- Realizar una evaluación económica que estime los beneficios relacionados a la aplicación de ofertas basadas en el modelo final.

4. MARCO CONCEPTUAL

4.1 DATA MINING (Minería de Datos)

Según Jiawes Han y Micheline Kamberen, en el libro “Data Mining: Concepts and Techniques”, en la actualidad existe una amplia disponibilidad a enormes cantidades de datos y surge la necesidad de convertirlos en información útil y conocimiento agregado para las empresas. El progreso constante de la tecnología de hardware en las últimas tres décadas ha dado lugar a grandes cantidades de computadores, equipamiento de recolección de datos y medios de almacenamiento para la gestión, recuperación y análisis de datos.

Los datos se pueden almacenar en diferentes tipos de bases de datos y repositorios de información. Como por ejemplo, en bases de datos relacionales, en almacenes de datos, en bases de datos transaccionales, en sistemas avanzados de bases de datos, en ficheros categóricos, en flujos de datos y en la World Wide Web. Esta abundancia de datos y el acceso a herramientas de análisis de datos, han hecho que las empresas tengan abundancia de datos pero pobreza en la obtención de información relevante. En consecuencia, las decisiones importantes no se basan en la información, sino en la intuición del que toma la decisión.

Las herramientas de minería de datos permiten descubrir patrones ocultos en grandes cantidades de datos, es decir, su aporte se basa en entender la información disponible y por consiguiente contribuir en gran medida a la estrategia del negocio.

La metodología KDD (Knowledge Discovery in Data)¹¹ consiste en una secuencia iterativa de pasos, donde los datos son procesados, integrados, seleccionados, transformados y aplicados a las técnicas de minería de datos con el fin de extraer información oculta. Además, se deben identificar los patrones realmente interesantes [1] (Anexo 9). Dentro de la metodología KDD, se desprende la metodología CRISP DM que consiste en un modelo jerárquico, el cual posee 4 niveles (Anexo 11). Esta metodología aporta con un mayor detalle en la comprensión del negocio. El modelo cuenta con un nivel superior, correspondiente a seis fases y cada una de ellas, conlleva ciertas tareas genéricas de segundo nivel. El tercer nivel, corresponde a tareas especializadas, donde se deben llevar a cabo las tareas genéricas, a partir de situaciones específicas. Y por último, el cuarto nivel, es un registro de las acciones, decisiones y de los resultados de la minería de datos [16] (Anexo 12).

Estas metodologías buscan integrar técnicas de múltiples disciplinas, tales como tecnologías de bases de datos y de almacenamiento de datos, estadísticas, aprendizaje automático, computación de alto rendimiento, reconocimiento de patrones, redes neuronales, visualización de datos, recuperación de información y análisis de datos temporales. Estas técnicas pueden tener dos enfoques: Modelos Descriptivos o Modelos Predictivos.

El presente trabajo se desarrolló a partir de bases de datos relacionales. Es decir, corresponde a una recolección de tablas y bases, donde a cada una de ellas, se le asigna un nombre único. Cada tabla está compuesta por un conjunto de atributos y por lo general poseen un gran conjunto de tuplas. A su vez, cada tupla es un objeto identificado por una clave única.

Las tareas de minería de datos se pueden clasificar en dos categorías: **Descriptivas** y **Predictivas** (Anexo 10). Las tareas **Descriptivas son Modelos No Supervisados** que caracterizan las propiedades generales de los datos, las medidas de error son difusas y existen muchas variables que describen distintos objetivos. Las tareas **Predictivas son Modelos Supervisados** que realizan inferencias en los datos con el fin de hacer predicciones, existe mayor claridad en medidas de errores y existe una variable objetivo a modelar [1].

¹¹ En español, Descubrimiento del Conocimiento a partir de los Datos

La minería de datos tiene el potencial de generar miles de patrones o reglas pero generalmente no todos son de gran interés. Para esto se deben establecer medidas objetivas, que se basan en la estructura de los patrones y las estadísticas.

4.1.1 MODELOS PREDICTIVOS

Las tareas predictivas son Modelos Supervisados que realizan inferencias en los datos con el fin de hacer predicciones, existe mayor claridad en medidas de errores y existe una variable objetivo a modelar.

Dentro de las tareas predictivas se distingue la **Clasificación** y la **Predicción**. La clasificación es el proceso de búsqueda de un modelo que describe y clasifica clases de datos, con el fin de poder utilizar el modelo para predecir categóricamente (discreto y sin orden) la clase de objeto cuya etiqueta es desconocida. En cambio, la predicción es el proceso de búsqueda de un modelo o funciones de valores continuos para predecir valores numéricos que faltan o no están disponibles dentro de los datos para extraer información de los datos [1].

Algunas de las técnicas o algoritmos asociados a modelos predictivos son:

- Regresiones
- Árboles de Decisión
- Redes Neuronales
- Serie de Tiempo
- Clasificaciones Bayesianos
- Lógica Difusa

El presente trabajo de título empleó árboles de decisión, regresiones logísticas y redes neuronales. Para los cuales se debió entrenar el sistema usando datos que contengan el comportamiento de los clientes, de tal forma de predecir el comportamiento del cliente usando patrones que caracterizan el uso de ciertos servicios.

Árbol de Decisión

Estos modelos aproximan funciones que toman valores discretos representando lo aprendido por medio de una estructura de árbol. Tienen la característica de ser robustos frente a datos con ruido y de ser expresiones disyuntivas, es decir, los árboles aprendidos se pueden representar como reglas intuitivas (if-then) [2].

Los Árboles de Clasificación funcionan de la siguiente manera. De N registros, se parte con L ($L \leq N$) etiquetados de la siguiente forma:

$$L = \{ (x_i, y_i), i = 1, 2, \dots, N; y_i \in \{1, 2, \dots, C\} \}$$

Donde (x_i, y_i) está descrito por un vector de atributos predictores x_i y una etiqueta de clase y_i perteneciente a alguna de las C clases (clasificación de los registros). En este caso, la clasificación corresponde si el cliente ha realizado up-selling o aún no lo ha hecho. El objetivo del árbol es construir un clasificador que, dado un nuevo ejemplo sin etiquetar caracterizado por el vector de atributos x (no incluido necesariamente en la etapa de entrenamiento), prediga la clase y a la que pertenece usando el conocimiento aprendido en el conjunto de datos inicial L [3].

La representación como árbol se caracteriza por tener nodos, ramas y nodos terminales (hojas). Cada **nodo** (no terminal) especifica un test de algún atributo de la instancia. Cada **rama** corresponde a un posible valor del atributo. Y cada **nodo terminal** indica la clase en la que se clasifica. Además, las instancias no vistas se clasifican recorriendo el árbol, es decir, pasando por los test de cada nodo, partiendo desde el nodo raíz hasta algún nodo terminal, el cual entrega su clasificación.

Dentro de los árboles de decisión se pueden identificar los **Árboles de Clasificación** y los **Árboles de Regresión**. Los árboles de clasificación se utilizan cuando la función objetivo toma valores discretos y los árboles de regresión se utilizan cuando la función objetivo toma valores continuos. Para este trabajo se ocupó árboles de clasificación, ya que el problema a resolver es si un cliente realiza o no realiza un up-selling a un Plan Multimedia.

Algunas de las ventajas de estos modelos es que son robustos a errores, tanto en la variable de respuesta como en los atributos, puede haber datos perdidos (missing values) para algunas instancias en los valores de los atributos y pueden trabajar con dominios complejos [2].

C5.0 Versión Comercial

El C5.0 [Quilan, 1997] es un algoritmo de inducción que busca maximizar la información relevante en base a la clasificación de los registros. Es la versión comercial del algoritmo C4.5, donde sus antecesores precisamente son los algoritmos C4.5 y ID3. El algoritmo C5.0 utiliza la estrategia “divide y reinarás” para generar el árbol de decisión inicial a partir de un conjunto de datos de entrenamiento. El método “divide y reinarás” realiza en cada paso una partición de los datos del nodo según una prueba realizada sobre el mejor atributo [4].

ID3¹² [Quilan, 1986] pertenece a la familia de los TDIDT¹³ que son modelos inductivos del Aprendizaje Automático, que aprenden a partir de ejemplos preclasificados. Para construir el árbol de decisión, a partir de un conjunto L de datos de entrenamiento, se deben definir las clases donde los datos se etiquetarán $\{C_1, C_2, C_3 \dots, C_k\}$ donde el objetivo es clasificar el conjunto L en subconjuntos de casos que tiendan hacia una colección de casos pertenecientes a una única clase [5]. ID3 utiliza una estrategia de búsqueda voraz por el espacio de posibles árboles de clasificación y construye el árbol de arriba hacia abajo, preguntándose qué atributo seleccionar como nodo raíz. Al ser voraz, esta técnica resuelve el problema a través de una mata heurística consistente en elegir la opción óptima en cada paso local con el fin de llegar a una solución óptima global. La metodología que ocupa es evaluar cada atributo para determinar cuán bien clasifica los ejemplos, luego se selecciona el mejor atributo como nodo, se abre el árbol para cada posible valor del atributo y los ejemplos se clasifican y colocan en los nodos apropiados. Se repite todo el proceso utilizando los ejemplos asociados con el nodo en el que se esté (siempre hacia delante, buscando los nodos no usados). El proceso finaliza cuando el árbol clasifica correctamente los ejemplos o cuando se han usado todos los atributos [2].

El algoritmo efectúa una selección de variables previa, denominada “pre pruning”, que consiste en efectuar un test de independencia entre cada variable explicativa X_i y la variable clase C .

El paso clave del modelo es cómo seleccionar el atributo. ID3 escoge la variable más efectiva usando la Ganancia de Información y la Reducción Esperada de Entropía (incertidumbre). Es decir, el atributo seleccionado será el que posea un mayor valor en la ganancia de información.

$$I(C, X_i) = H(C) - H(C/X_i) \quad \text{"Ganancia de Información"}$$

$$H(C) = - \sum_c p(c) \log_2 p(c); \quad H(C/X) = - \sum_c \sum_x p(x, c) \log_2 p(c / x)$$

Donde $I(C, X_i)$ es la información entregada por el vector de atributos X_i para predecir la clase a la que corresponde, $H(C)$ es la entropía o incertidumbre de la clasificación, C es la etiqueta de clasificación y X_i es el atributo i [2].

El sistema tiene algunas desventajas. Al ser voraz, puede que conduzca a una solución óptima local en vez de global. Por usar propiedades estadísticas de todos los ejemplos (ganancia de información) la búsqueda es menos sensible a errores en los datos, es decir, las hojas tienen la necesidad de esperar que todas las etiquetas sean

¹² En inglés, Induction Decision Trees

¹³ En inglés, Top Down Induction Trees

iguales. La complejidad crece linealmente con el número de instancias de entrenamiento y exponencialmente con el número de atributos. Se sobre ajusta el árbol a medida que se añaden niveles, ya que las hipótesis se refinan tanto que describen muy bien los ejemplos utilizados en el aprendizaje. Y por último, matemáticamente se demuestra que favorece la elección de variables con mayor número de valores [2].

De esta manera aparecen varios aspectos débiles del modelo:

- Cuánto hacer crecer el árbol (problemas de sobre ajuste).
- La medida de selección de los atributos es la adecuada.
- Cómo manejar atributos continuos.
- Cómo manejar datos de entrenamiento con valores perdidos [2].

El modelo C4.5 [Quilan, 1993] soluciona muchas de estas limitaciones y genera un árbol de decisión mediante particiones recursivas basado en la estrategia de profundidad-primero (depth-first). El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información.

Para solucionar el problema de sobre ajuste existen dos grupos de técnicas que tratan de simplificar el árbol:

- Pre Poda: Interrumpir el crecimiento del árbol antes de que alcance el punto en que clasifica perfectamente los ejemplos de entrenamiento.
- Post Poda: Permitir que sobre ajuste los datos y después podar reemplazando sub árboles por una hoja.

De esta manera los árboles son menos frondosos, ya que cada hoja cubre una distribución de clases y no una clase en particular.

Generalmente se utiliza la incorporación de post poda de reglas, donde se generan las reglas (una por camino) y se eliminan las pre condiciones siempre y cuando se mejore o se iguale el error.

Se basa en la utilización del criterio de proporción de ganancia, Ratio de Ganancia, que penaliza los atributos con mayor cantidad de valores y con una distribución uniforme [2].

$$\frac{I(C, X_i)}{H(X_i)} \quad \text{"Gain Ratio"}$$

Por otra parte, el modelo permite trabajar con valores continuos, separando los posibles resultados en dos ramas ($A_i \geq n \wedge A_i \leq n$). Y por último, los atributos con

valores perdidos se imputan, es decir, se eliminan las instancias incompletas. Esto último se realiza asignándole la moda o la distribución de probabilidad [5].

Por último, el algoritmo C5.0 posee una metodología similar a la del el algoritmo C4.5, con la gran diferencia de que el código es cerrado. Las diferencias más preponderantes dentro de estos dos modelos, a favor del algoritmo C5.0, son:

- Conjunto de reglas más eficientes.
- Algoritmo mucho más rápido.
- Menor uso de memoria.
- Se pueden utilizar árboles de decisión muchos más pequeños y obteniendo resultados similares.
- Creación de “bootstrapping” (método de re muestreo para aproximar sesgo o varianza) para generar mayor precisión.
- Creación de ponderadores de peso para nivelar cada tipo de error.
- Mejor tolerancia al ruido y datos nulos [7].

Existen otros algoritmos para árboles, como CART [Breiman et al., 1984] [3] y CHAID [Kass, 1984] (CHi-square Automatic Interaction Detection) [6], que en este trabajo fue necesario utilizarlos ya que otros modelos entregaban mejores resultados. En la figura 1, se comparan los tres métodos de árboles de decisión nombrados con anterioridad para tener claridad en las características que posee cada uno.

Figura 1: Comparativa entre algoritmos de aprendizaje para árboles de decisión.

Algoritmos	Variables de entrada	Variable de salida	Tipo de predicción	Ramas por división	Criterio de división
CHAID	Categorica/Numérica	Categorica/Numérica	Clasificación/Regresión	≥ 2	Chi-cuadrado/F
CART	Categorica/Numérica	Categorica/Numérica	Clasificación/Regresión	$= 2$	GINI/Desviación cuadrática mínima
C4.5/C5.0	Categorica/Numérica	Categorica	Clasificación	≥ 2	Gain Ratio

Fuente: Elaboración Propia.

Redes Neuronales

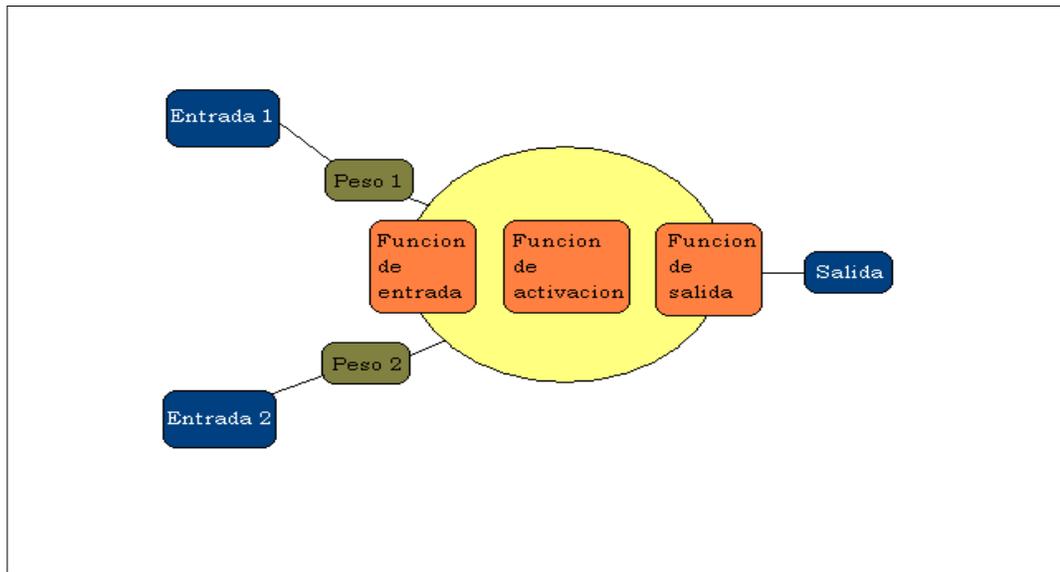
Las Redes Neuronales son modelos artificiales y simplificados del cerebro humano, que es el ejemplo más perfecto del que disponemos para crear un sistema que sea capaz de adquirir conocimiento a través de la experiencia. La unidad básica del procesamiento de la información está inspirada en la célula fundamental del sistema nervioso humano: *la neurona*.

Algunas de las ventajas de estos modelos son [8]:

- Aprendizaje Adaptativo: Capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.
- Auto Organización: Puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.
- Tolerancia a fallos: La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.
- Operación en tiempo real: Los cómputos neuronales pueden ser realizados en paralelo.

En la figura 2 se pueden ver los elementos básicos que constituyen una neurona.

Figura 2: Neurona con dos entradas y una salida.



Fuente: Redes Neuronales, Conceptos Básicos y Aplicaciones [8].

El primer elemento que constituye una red neuronal es la **Función de Entrada**, donde la neurona trata a muchos valores de entrada como si fuera uno solo, esto recibe el nombre de *entrada global*. De esta manera se deben combinar las entradas $(in_{i1}, in_{i2}, \dots, in_{in})$ dentro de la entrada global gin_i . Entonces, la función de entrada se puede describir como sigue:

$$gin_i = (in_{i1} \cdot w_{i1}) \blacksquare (in_{i2} \cdot w_{i2}) \blacksquare \dots \blacksquare (in_{in} \cdot w_{in})$$

Donde \blacksquare representa el operador apropiado (por ejemplo: máximo, sumatoria, pitagórica, etc.), n es el número de entradas a la neurona N_i y w_i es el peso correspondiente a la entrada [8].

La neurona artificial puede estar activa o inactiva, es decir que tiene un estado de activación. La **Función de Activación** calcula el estado de actividad de una neurona, transformando la entrada global en un valor de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1). La función de activación es una función de la entrada global (gin_i) menos el umbral (θ_i) . Dentro de las funciones de activación más ocupadas se encuentran la función lineal, la función sigmoidea y la función tangente hiperbólica [8].

El último componente que una neurona necesita es la **Función de Salida**. El valor resultante de esta función es la salida de la neurona i (out_i), por ende, la función de salida determina que valor se transfiera a las neuronas vinculadas. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona subsiguiente. Los valores de salida están comprendidos en el rango [0,1] o

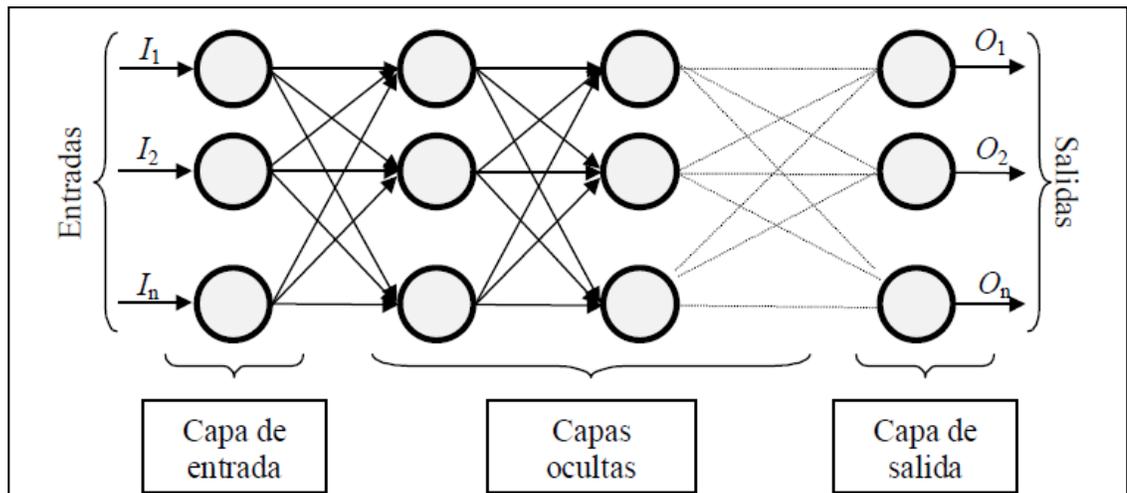
$[-1,1]$. También pueden ser binarios $\{0,1\}$ o $\{-1,1\}$. Dentro de las funciones de salida más comunes se tiene [8]:

- Ninguna: Donde la función de salida es la misma que la función de entrada.
- Binaria: Donde toma el valor 1 si se activa la neurona o 0 si es que ocurre lo contrario.

Existen tres tipos de capas o agrupaciones de neuronas. La capa de entrada, las capas ocultas y la capa de salida. La capa de entrada recibe la información proveniente de las fuentes externas de la red. Las capas ocultas son internas a la red y no tienen contacto directo con el entorno exterior (puede haber más de una). Y la capa de salida transfiere la información de la red hacia el exterior.

En la figura 3 se muestran los componentes de una red neuronal

Figura 3: Red Neuronal totalmente conectada.



Fuente: Redes Neuronales, Conceptos Básicos y Aplicaciones [8].

Una red neuronal debe aprender a calcular la salida correcta para cada vector de entrada en el conjunto de ejemplos. Este proceso de aprendizaje se denomina proceso de entrenamiento. El conjunto de datos sobre el cual este proceso se basa es llamado conjunto de datos de entrenamiento. El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios que se producen durante el mismo se reducen a la destrucción, modificación, y creación de conexiones entre las neuronas. Durante el proceso de aprendizaje, los pesos sufren modificaciones, por lo tanto, se puede afirmar que este proceso ha terminado cuando los valores de los pesos permanecen estables ($\frac{dw_{ij}}{dt} = 0$).

Para determinar cuándo se detendrá el proceso de aprendizaje, es necesario establecer una *condición de detención*. Normalmente se detiene cuando el cálculo del error cuadrado sobre todos los ejemplos de entrenamiento ha alcanzado un mínimo o cuando para cada uno de los ejemplos dados, el error observado está por debajo de un determinado umbral. Otra condición de detención puede ser cuando un cierto número de ciclos y/o pasos de entrenamiento hayan sido completamente corridos. Este modelo puede trabajar con variables continuas o variables discretas [8].

Regresión Logística

El modelo de regresión logística se utiliza cuando interesa pronosticar la probabilidad de que ocurra o no un suceso determinado. Donde es posible estudiar el impacto que tiene cada una de las variables explicativas en la probabilidad de que ocurra el suceso en estudio.

El análisis de la regresión logística es una herramienta muy flexible en cuanto a la naturaleza de las variables explicativas, pues estas pueden ser discretas o continuas. La variable dependiente ficticia es binaria.

El modelo inicial es una regresión lineal a partir de las variables explicativas [9].

$$y = p = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + u$$

Donde se puede notar que el valor entregado no está acotado dentro del intervalo $[0,1]$, y, por lo tanto, ya no va a representar una probabilidad. Una forma de garantizar que los valores pronosticados estén en el intervalo $[0,1]$ es considerar la siguiente transformación.

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + u$$

Con el fin de simplificar la notación, se define z:

$$z = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n$$

Por lo tanto, el modelo se puede representar como:

$$\ln\left(\frac{p}{1-p}\right) = z + u$$

Donde p es la probabilidad de que ocurra el suceso de estudio. Aplicando la función de distribución logística.

$$F(x) = \frac{e^x}{1 + e^x}$$

Por tanto, podemos reescribir el modelo de la siguiente forma:

$$p = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}} = F(z) = F(x \cdot b)$$

De donde se deduce que el modelo de regresión logística es, en principio, un modelo de regresión no lineal que se puede transformar a escala logarítmica. Es así, como existen dos formas de expresar el modelo.

$$\ln(p) - \ln(1 - p) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n \quad \text{"Logit"}$$

$$\frac{p}{1 - p} = e^{b_0} \cdot e^{b_1 \cdot x_1} \cdot e^{b_2 \cdot x_2} \cdot \dots \cdot e^{b_n \cdot x_n} \quad \text{"Odds Ratio"}$$

Si bien en la regresión lineal la interpretación de los coeficientes de regresión es simple e intuitiva (b_k es el incremento producido en la variable dependiente por un incremento unitario en la variable x_k), en la regresión logística no es tan intuitiva, al depender tanto del valor de x_k , donde se produzca el incremento, como del valor del resto de las variables, pues la pendiente de la curva de regresión varía. Para mejorar la interpretación, se ocupa el *Odds Ratio* como el cociente de probabilidades entre que ocurra un suceso respecto de que no ocurra. Entonces, los coeficientes B indican el incremento de la probabilidad de que ocurra suceso en escala logarítmica [9].

En síntesis, en figura 4 se presenta una comparativa entre las técnicas de minería de datos que son utilizadas en el proyecto.

Figura 4: Comparativa entre las técnicas de Data Mining utilizadas

Comparación de propiedades generales entre los modelos			
	Regresión Logística	Redes Neuronales Artificiales	Árboles
Accesibilidad	++	-	-
Tipos de variables	++	++	++
Selección de variables	++	-	++
Importancia de variables	++	+	++
Interrelación de variables	+	+++	++
Modelo	+	+++	++
Interpretación de resultados	++	-	+++

Se ha graduado desde (-), peor, a (+++), mejor, las propiedades evaluadas

Fuente: http://scielo.isciii.es/scielo.php?pid=S0213-91112008000100013&script=sci_arttext

A continuación se detalla cada una de las propiedades de los modelos que fueron comparadas.

- **Accesibilidad:** Los paquetes estadísticos de uso habitual incluyen la RLM (Regresión Logística). Hay infinidad de publicaciones en medicina que emplean esta metodología. Los modelos basados en RNA (Redes Neuronales Artificiales) y CART no son tan accesibles, ya que precisan programas específicos de menor difusión.

- **Tipos de variables:** Posibilidad de utilizar variables continuas o discretas. Los 3 modelos ofrecen esa posibilidad.

- **Selección de variables según la importancia y la aportación de información:** En RL y CART se hace de forma automática, y en RNA debe hacerse de forma «manual» (aunque hay otros programas que incorporan esta posibilidad).

- **Modelo de asignación de probabilidad.** El modelo RL es paramétrico (asignación según modelo logístico). El basado en CART es no paramétrico y las redes son el paradigma de aproximación universal de funciones.

- **Interrelación de variables:** En RL deben incluirse en el modelo, CART analiza las interrelaciones en sus reglas de decisión y en RNA pueden tenerse en cuenta todas las posibilidades.

- **Interpretación de resultados:** Aquí la ventaja está a favor del CART, es menor en RL y, en el caso de las RNA, son una caja negra en la interpretación de sus parámetros.

- **Los modelos no son excluyentes:** ante un mismo problema deben barajarse distintas metodologías que pueden colaborar en la búsqueda de un resultado óptimo.

Como conclusión, podemos afirmar que los modelos basados en árboles de decisión ofrecen como ventaja una simplicidad en su utilización e interpretación ya que sus reglas de decisión no necesitan procesos matemáticos para ser interpretadas [10].

5. METODOLOGÍA

El desarrollo de este proyecto se basa en dos grandes metodologías como son, KDD (Knowledge Discovery in Data)¹⁴ y CRISP DM (Cross Industry Standard Process for Data Mining)¹⁵. Ambas metodologías buscan identificar interesantes patrones ocultos en grandes conjuntos de datos.

La metodología KDD (Knowledge Discovery in Data)¹⁶ consiste en una secuencia iterativa de los siguientes pasos [1] (Anexo 9):

1. **Limpieza de datos:** Eliminar ruidos y los datos inconsistentes.

¹⁴ En español, Descubrimiento del Conocimiento a partir de los Datos

¹⁵ En español, Metodología para el Desarrollo de Proyectos en Minería de Datos.

¹⁶ En español, Descubrimiento del Conocimiento a partir de los Datos.

2. **Integración de datos:** Múltiples fuentes de datos pueden ser combinados.
3. **Selección de datos:** Se seleccionan los datos relevantes para la tarea de análisis.
4. **Transformación de datos:** Los datos se transforman o consolidan mediante la realización de resúmenes o de agregación.
5. **Minería de datos:** Proceso esencial donde los métodos son aplicados con el fin de extraer patrones ocultos.
6. **Modelo de evaluación:** Se identifican los patrones realmente interesantes representando el conocimiento basado en algunas de las medidas de grado de interés.
7. **Presentación del conocimiento:** Las técnicas de visualización y de representación del conocimiento se utilizan para presentar los resultados extraídos al usuario.

Dentro de la metodología KDD, se desprende la metodología CRISP DM que consiste en un modelo jerárquico, el cual posee 4 niveles (Anexo 11). Esta metodología aporta con un mayor detalle en la comprensión del negocio. El modelo cuenta con un nivel superior, correspondiente a seis fases y cada una de estas, conlleva ciertas tareas genéricas de segundo nivel.

El modelo cuenta con 6 fases [16] (Anexo 12):

1. **Comprensión del negocio:** Comprensión de los objetivos del negocio, para así poder definir el objetivo de la minería de datos.
2. **Comprensión de los datos:** Colección de datos iniciales, familiarización con los primeros datos, identificar la calidad de los datos, describir conocimientos en los datos y formulación de hipótesis.
3. **Preparación de los datos:** Construcción de los datos finales, selección de las tablas, registros, y atributos, así como la transformación y limpieza de los datos para la modelación.
4. **Modelo:** Técnicas de modelación sobre los datos.
5. **Evaluación:** Evaluar el modelo a partir de tablas de confusión y revisar los pasos ejecutados para comparar el modelo obtenido con los objetivos del negocio.
6. **Desarrollo:** El conocimiento obtenido debe ser organizado y presentado a los clientes del negocio.

A partir de las dos metodologías, mencionadas anteriormente, se buscó generar un proceso que mezclara a ambas y rescatará los detalles más importantes de cada una de ellas. Cabe destacar la metodología generada no es la suma de todas las fases, ya que en ambas metodologías hay fases que se repiten.

1. Comprensión del negocio

La etapa inicial busca comprender, desde la perspectiva del negocio, lo que la empresa quiere lograr. Se desea investigar la situación actual de la industria para determinar los objetivos del análisis de los datos y el desarrollo del plan de trabajo. A partir de los objetivos del negocio detectados es posible definir los objetivos de la minería de datos. De esta manera se podrá transformar este conocimiento en la definición de un problema de Minería de Datos.

2. Identificación y planteamiento del problema a investigar

En esta fase se busca definir los objetivos, la justificación, los antecedentes generales, la descripción y los alcances del proyecto. De tal manera de ligar los objetivos del negocio con los objetivos de minería de datos, es decir promover la experiencia multimedia y aumentar el ingreso promedio por usuario con predecir cuales son los clientes más propensos a comprar Planes Multimedia.

3. Establecer e investigar el marco teórico correspondiente a la solución del problema

Esta etapa estará orientada a investigar y procesar la teoría necesaria para el desarrollo del proyecto. Es decir, investigar cómo funcionan los algoritmos de minería de datos. En este caso, Árbol de Decisión, Red Neuronal y Regresiones Logística.

4. Caracterización del problema y su respectiva solución

Esta fase corresponde a plantear el problema y diseñar la metodología apropiada para resolver el problema.

5. Recopilación de bases de datos relacionadas a la resolución del problema

Esta fase está orientada a la recolección de fuentes y bases de datos que contengan información transaccional, socio-demográfica y financiera de los clientes. Lo importante de esta fase, es que se deben elegir las variables o datos más apropiados para alcanzar el objetivo del proyecto. Para ello es fundamental los comentarios de expertos y literatura con respecto al negocio.

6. Comprensión de los datos

Se debe realizar un resumen descriptivo de los datos, donde es importante obtener medidas de tendencia central, medidas de dispersión, gráficos y descripción de los datos. Es fundamental identificar los problemas de calidad de los datos y descubrir los primeros conocimientos para formular una hipótesis en cuanto a la información oculta.

7. Preparación de los datos

Cubre todas las actividades necesarias para construir el conjunto final de los datos. Se destaca selección de tablas, registros y atributos. Se debe realizar la limpieza de los datos donde se eliminan ruido y datos inconsistentes. Además, se deben integrar los datos, donde múltiples fuentes de datos pueden ser combinadas. Por otra parte, es necesario recuperar los datos relevantes para la tarea de análisis de la base de datos y extraer los datos finales sobre los cuales se aplicara el modelo de Minería de Datos. Por último, los datos se deben transformar en formas apropiadas para la explotación. Además, se busca reducir dimensión de los datos y la numerosidad de ellos. Es posible que en esta fase se creen variables a partir de las existentes, de tal manera de poder tener variables que aporten en la predicción del modelo.

8. Minería de datos

Proceso esencial donde se aplican las técnicas de Data Mining, con el fin de extraer patrones ocultos. Las técnicas aplicadas fueron Redes Neuronales, Árboles de Decisión y Regresiones Logísticas. Se compararan las técnicas para obtener la solidez necesaria en la predicción. La elección de estas técnicas se debe a su poder explicativo y su fácil entendimiento para poder crear estrategias comerciales.

9. Modelo de evaluación

Se busca identificar patrones interesantes basados en algunas de las medidas de grado de interés, ya sea para el negocio como para el objetivo de la minería de datos. Además, se compararan los objetivos del negocio con los resultados obtenidos.

10. Presentación del conocimiento y conclusiones

El conocimiento adquirido deberá ser organizado y presentado en un modo sencillo. Además, como conclusión se debe entregar el dimensionamiento del impacto del proyecto, es decir, estimar una evaluación económica que entregue la ganancia de aplicar el conocimiento adquirido en los modelos.

6. ALCANCES

El modelo de propensión de compra ayudará a crear ofertas focalizadas para clientes con baja vinculación dentro de la compañía o con altas probabilidades de realizar up-selling a Planes Multimedia.

Los Planes de Voz Multimedia (Planes de Voz + Mensajes + Internet Móvil) que se consideraron dentro del proyecto son 10 planes con distintas características

(minutos a todos destino) y 4 planes Full Red con distintas características (minutos a Entel y Red Fija, y por separado, minutos a otras compañías).

El modelo de propensión de compra no dará ningún resultado sobre la reducción de las migraciones de los clientes ni información sobre la experiencia del cliente en las tiendas.

7. RESULTADOS ESPERADOS

Se desea obtener un modelo de propensión de compra para incrementar los ingresos promedios por usuario y promover la experiencia multimedia dentro de los clientes persona de Entel. Se espera poder generar ofertas focalizadas a los clientes con los patrones y características aprendidas con los modelos de predicción.

Se busca generar, evaluar y seleccionar él o los modelos predictivos con mayor calidad de predicción. Con los resultados de los modelos, se podrá identificar y caracterizar a los clientes propensos a comprar las ofertas focalizadas con respecto a los productos y servicios target.

Y por último, es fundamental obtener una evaluación económica sobre los beneficios que traería consigo la implementación de las ofertas focalizadas.

8. DESARROLLO DE LA METODOLOGÍA

En este capítulo se desarrolló la segunda parte de la metodología, que busca alinear los objetivos del proyecto con la solución del problema. Fue así como el proceso abarca la recopilación de las bases de datos, además de la comprensión, preparación, selección y transformación de los datos y por último la modelación de estos para predecir el up-selling a Planes Multimedia.

8.1 RECOPIACIÓN DE LAS BASES DE DATOS

Esta etapa es fundamental dentro del desarrollo del proyecto, ya que las variables descriptivas que caracterizan a los clientes son la base de la predicción. En este sentido, la recopilación de los datos debe ser la adecuada, pues deben estar incluidas las variables que los expertos del negocio identifican como explicativas para realizar up-selling y además posibles variables que puedan entregar información útil sobre patrones ocultos.

La recopilación de las bases de datos se realizó mediante dos canales. El primero, fue a través de reuniones con el Área de Ventas, el Área de Marketing y el Área de Gestión del Valor de los Clientes, donde se pidieron recomendaciones de variables relevantes para el estudio. Y el segundo canal, fue a través de bases de datos entregadas por el Área de Data Warehouse de la compañía, según recomendaciones de expertos y por intuición del memorista.

Las tablas de datos seleccionadas proveen información demográfica, geográfica, transaccional, tráfico de minutos (emisión y recepción), tráfico de mensajería (emisión y recepción), tráfico de datos (Internet), equipos terminales (celulares y otros), facturación, recargas, bolsas de servicios y riesgo financiero.

A continuación se nombran y se describen los campos pertenecientes a cada una de las tablas. En el Anexo 13 se puede consultar con más detalle.

La tabla **BSCS**, contiene información general de los clientes Pos-Pago de la compañía. Dentro de los campos más relevantes se encuentran: móvil, rut, plan de VOZ, nombre del cliente, estado actual del servicio (activo o suspendido), segmento (asignado por la compañía), comuna, ciudad, fecha ingreso, entre otros.

La tabla **BUIC**, contiene información demográfica y geográfica de los clientes. Sus principales campos son: número móvil, descripción del plan, fecha de ingreso, rut, tipo de cliente (usuario o dueño), nombre del cliente, género, fecha de nacimiento, estado civil, dirección del cliente, email, nivel de estudio, actividad, profesión, hijos, fecha desde la última modificación de estado, ciudad, comuna y sector económico.

La tabla **Terminales**, especifica el número móvil, la marca, el modelo y la tecnología del equipo.

El **Tráfico Mensual de Minutos**, se descompone en número móvil, minutos emitidos y minutos recibidos.

La tabla **Tráfico de Mensajería** contiene información sobre el tráfico de mensajes de texto y mensajes multimedia. En cada tipo de mensaje, se posee el tráfico de emisión y de recepción del cliente.

El **Tráfico Mensual de Datos** contiene la información de kilo bytes descargados mensualmente por clientes que hayan utilizado servicios de datos.

Para analizar la facturación mensual de los clientes se trabajó con tres tablas que se detallan a continuación.

La tabla **Facturación** detalla la cuenta mensual asociada a los servicios utilizados por el cliente y cobrados en su factura. En particular, posee los siguientes campos:

número móvil, descripción del plan, concepto de facturación, servicio traficado, unidad del servicio traficado, cantidad de tráfico y el monto facturado por el servicio.

La segunda tabla contiene información sobre las **Recargas** que los clientes realizan mensualmente (para planes Cuenta Controlada). La información entregada por esta tabla es la siguiente: número móvil, cantidad de recargas, monto de la recarga, bono de recarga (saldo adicional aplicado).

Y la tercera tabla corresponde a las **Bolsas** de servicios (mensajes, Internet, minutos, etc.) contratadas por los clientes. Los campos que contiene esta son: móvil, tipo de bolsa, monto de la bolsa y la cantidad del servicio contratados.

Además, se analiza el **Riesgo Financiero** de los clientes. Donde se analizan indicadores de atraso en los pagos y actividades de cobranza que realiza la compañía cuando un cliente ha dejado de pagar en la fecha que corresponde.

La tabla **Atraso Promedio** contiene información sobre los días que el cliente se atrasa en su pago establecido (promedio de los últimos seis meses). Los campos que posee son los siguientes: Rut del cliente, suma de los días en que el cliente se ha atrasado los últimos seis meses y el atraso promedio de los últimos seis meses.

La tabla **Acciones de Cobranza** posee información de acciones de cobranza que realiza la empresa a los clientes. Los campos que contiene esta última son: número de contrato, la acción de cobranza que se realizó y el Rut del cliente.

Por último, los clientes pueden acceder a equipos terminales (celulares) a través de la modalidad de arriendo, donde el equipo es propiedad de Entel durante 18 meses. Al momento que se cancela la cuota de arriendo número 18 pasa a ser propiedad del cliente. Según esta modalidad de pago los clientes optan por descuentos que se aplican en la factura mientras se tenga vigente el contrato de suministro. La tabla que entrega esta información tiene por nombre **ACOC** y posee los siguientes campos: Rut del cliente, número de contrato, estado de pago, cuotas totales, cuotas pagadas, valor de la cuota, valor de la opción, entre otras.

En general, las tablas contenían la información de todos los clientes del Mercado Personas de la compañía (aprox. 2,5 MM de clientes). Sin embargo, los clientes target de este estudio pertenecer al Mercado Personas con Planes Suscripción o Cuenta Controlada, que aproximadamente tiene un total de 1,3 MM de clientes.

Para gestionar las bases de datos se utilizó la herramienta PostgreSQL, caracterizada por ser OpenSource, tener sistemas estables y de alto rendimiento para ambientes de alto volumen y además posee gran flexibilidad frente a diversos sistemas operativos.

8.2 COMPRESIÓN DE LOS DATOS

Dado que el estudio comprende 1,3 MM de clientes, la carga y la gestión de estos tomó bastante tiempo de ejecución. Para solucionar este problema y agilizar el avance, se decidió trabajar con el 10% de los datos, número que corresponde a 130.000 clientes. Esta cantidad de datos siguió siendo representativa en la tendencia y en las medidas de las variables, por lo que los resultados finales se pueden extrapolar para toda la muestra con un sesgo mínimo.

La llave primaria de las tablas fue el número móvil del cliente. El Rut no fue elegido como llave primaria debido a que un cliente puede tener asociado más de un número móvil a su Rut, de esta forma se mezclarían los comportamientos de las distintas líneas que posee el cliente.

Con respecto a los atributos que caracterizan a los planes, se pudo consolidar la información de gran parte de ellos, sobre todo, la información de los planes comercialmente vigentes. Para el caso de los planes no vigentes (planes que no se encuentran en la oferta comercial pero aún existen clientes que los tienen contratados) no fue posible contar con toda su información. Sin embargo, en la totalidad de los casos se pudo levantar la información más relevante, como ejemplo el cargo fijo, los minutos a todo destino, los minutos a Entel y Red Fija, los minutos a otras compañías, los mensajes de texto, los mensajes multimedia y la forma del plan (Suscripción o Cuenta Controlada).

Para las variables demográficas se identificaron grandes problemas en la calidad de los datos. Esto se debe a que existen pocas variables pobladas, dado que son los mismos clientes los que entregan esta información. En este sentido, las únicas variables que poseían más de un 90% de los datos fueron género, región, grupo, estado, años en la compañía, edad y cantidad de líneas. En otra categoría, variables que poseían más de un 60% de los datos fueron estado civil, actividad y sector económico. Y por último, las variables consideradas en estado crítico por su calidad fueron nivel de estudio, profesión y cantidad de hijos.

Para el caso de los equipo terminales que están asociado a los números móviles, la calidad de los datos fue bastante buena. El 90% de los números móviles tenían asociado su respectivo equipo terminal.

Las variables transaccionales se encontraron en buena calidad, salvo un 4,36% de los valores que se encontraban sin poblar dado que algunos clientes no habían realizado transacciones.

Para los datos correspondientes a Facturación se tuvo que realizar un proceso de limpieza de datos, ya que para la mayoría de los Planes Multimedia el valor de la facturación traía involucrado un descuento de \$16.000.

Se pudo identificar un desfase de datos de ciertos clientes debido al prorrateo del consumo. Es decir, se identificaron clientes que contrataban un plan al final del mes, por lo tanto sus datos transaccionales solo reflejaban una proporción de su consumo mensual. El problema fue que no se tenían los datos para identificar en qué fecha del mes los clientes habían cambiado de plan. La solución se basó en definir ciertos criterios para detectar outliers.

En relación a las variables de riesgo financiero la calidad de los datos se encontró en condiciones óptimas. Por lo tanto, medidas de limpieza o procesamiento de datos no fueron necesarios.

En el caso de la cuota de arriendo del equipo terminal (ACOC), la calidad de los datos fue bastante mala. No se contaba con el 44,33% de los datos. Más aún, la única información disponible fue el acumulado del mes de Noviembre del 2011, es decir, no corresponde al periodo de estudio. La explicación de esta problemática se basa en que la empresa no lleva un registro histórico de esta variable, de tal manera que la información incorporada a los datos no es la adecuada. Sin embargo se decidió agregarla, para ver su comportamiento, ya que por recomendaciones de expertos, esta variable es relevante al momento de contratar un Plan Multimedia.

En conclusión, se estandarizaron los datos y se excluyeron casos que ensuciaban el análisis.

8.3 PREPARACIÓN DE LOS DATOS

El objetivo de esta etapa es construir el conjunto final de datos, de tal manera de poder obtener un conjunto sin sesgos ni errores en su conformación. Para conseguir esto, se obtuvieron medidas de tendencia central, medidas de dispersión, gráficos y descripción de los datos. Del mismo modo, al realizar estos análisis fue posible identificar, con mayor precisión, problemas en la calidad de los datos y además formular las primeras hipótesis sobre la propensión a contratar un Plan Multimedia.

Se utilizaron datos históricos desde Octubre del 2010 hasta Septiembre del 2011, por lo que el estudio toma datos con un año de historia. Cabe destacar, que al momento de hacer el requerimiento de los datos, se solicitaron dos años de historia, pero por motivos de política de la empresa, se manejan 13 meses móviles de

información disponibles a los usuarios. La historia se guarda en cintas, que se deben descriptar, pero es un proceso muy costoso para la compañía.

El primer filtro realizado fue seleccionar a clientes que se encuentren con un estado activo o suspendido en la compañía, es decir no se consideró a clientes que la empresa haya calificado como inactivos por deudas. Luego se filtró por segmento de clientes. Estos segmentos están definidos por la compañía y se muestran en la figura 5.

Figura 5: Segmentos de clientes en estudio.

Grupo Cliente	Glosa Grupo Clientes	Grupo 2	Mercado
1	Personas	Personas	Personas
10	Empleados de empresas	Personas	Personas
17	Clientes preferenciales	VIP	Personas
22	Control integral	Personas	Personas

Fuente: Elaboración Propia.

La elección de estos segmentos fue recomendada por expertos de la compañía. Se dejaron fuera de estudio grupos restrictivos para acceder a planes de mayor valor y además, se dejaron fuera a grupos de clientes que no tiene la toma de decisión al momento de elegir los planes.

El total de clientes del segmento target, después de aplicar el filtro anteriormente explicado, se redujo a 1.000.256 clientes. Esta cantidad de registros es bastante elevada para poder gestionar y procesar de manera rápida y eficiente. Por esta razón, como se comento en capítulos anteriores, se decidió extraer una muestra aleatoria del 10% para realizar el proyecto. Por lo tanto, la cantidad de registros finales termino siendo 100.256 clientes.

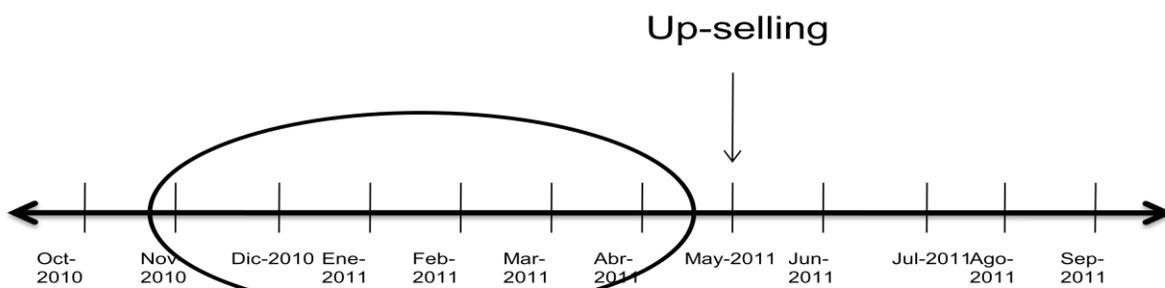
A continuación se detalla una serie de filtros que se aplicaron para la obtención del conjunto final de datos:

- Se tomaron solo aquellos clientes que han contratado un Plan Multimedia con más de dos meses de permanencia, dado que se busca encontrar perfiles de clientes que se hayan mantenido en el tiempo con el plan y no clientes que duraron sólo un mes o dos meses. Es decir, la idea es buscar clientes fieles al up-selling y que no lo hayan contratado solo por una promoción o descuento.

- No se consideraron clientes que hayan realizado up-selling antes del periodo de estudio ni clientes nuevos que hayan accedido a Planes Multimedia.

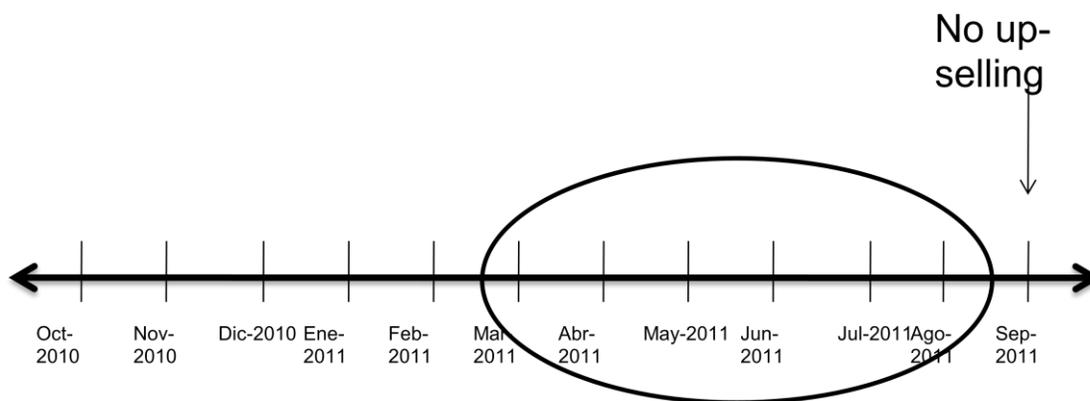
- Se analizaron todas las variables, con una historia de 6 meses, desde la fecha de contratación del Plan Multimedia, si fuese el caso de un cliente up-seller (Figura 6). Para el caso de los clientes que aún no realizan el up-selling se analizaron todas las variables, con una historia de 6 meses, desde la Septiembre del 2011 hacia atrás (Figura 7). Esto debido a que se debe comparar el comportamiento de los up-sellers con los clientes que aún no lo han realizado.

Figura 6: Periodos de estudio para clientes no up-seller.



Fuente: Elaboración Propia.

Figura 7: Periodos de estudio para clientes up-seller



Fuente: Elaboración Propia.

- Se consideran dentro del estudio sólo clientes que, por lo menos, posean historia de 4 meses continuos antes del periodo en estudio. Para el caso de los up-sellers por lo menos cuatro meses antes a su fecha de cambio de plan y para el caso de los que no han hecho up-selling por lo menos cuatro meses antes de Septiembre del 2011.

La idea de realizar los filtros, anteriormente expuestos, es generar un modelo de propensión de compra que se base en aprender del comportamiento y de las

características de los clientes que han realizado up-selling a un Plan Multimedia. La manera de compatibilizar y estandarizar las variables de diferentes tiempos cronológicos entre los clientes, es analizar el comportamiento durante el periodo en que ninguno tuviera un plan multimedia, de esta manera se puede identificar las similitudes previo al up-selling. De esta manera, es posible identificar patrones ocultos y poder predecir que un cliente que aún no realiza up-selling tenga altas probabilidades de hacerlo ya que tiene un comportamiento similar a los up-sellers. En palabras simples, los análisis futuros y los modelos predictivos buscan aprender se los clientes que se han cambiado a un Plan Multimedia, para identificar rasgos similares en los que se han mantenido con planes antiguos y poder ofrecer, comercialmente, el salto de plan.

Como existían algunos problemas de calidad en los datos, lo primero que se realizó fue una limpieza de registros. Se eliminaron los registros que tuvieran menos del 60% de los datos. Para el caso de registros que tuvieran valores perdidos hasta un 40% se decidió por rellenar con el promedio en el caso de valores numéricos y en el caso de valores categóricos se creó una nueva categoría que identifica que la variable no está informada dentro de los datos (categoría “-1”).

Para el caso de valores extremos, se definió que un outlier correspondía a un valor que se encontrara fuera del promedio más menos dos desviaciones estándar. De esta manera se identificaron los outliers y se reemplazaron por la media aritmética. Cabe destacar que fueron pocos los problemas de outlier, pero para el caso de los valores perdidos se tuvo que realizar más esfuerzos por depurar la información.

Uno de los pasos más relevantes dentro de la confección del conjunto final de datos, fue la integración de todas las tablas mencionadas en el capítulo 8.1. Para ello, fue necesario estandarizar todas las variables de tal manera que al entrecruzar las tablas no hubiera problemas de compatibilización (en especial se tuvo bastante precaución con las llaves primarias).

En conclusión, los atributos principales dentro del conjunto final de datos fueron:

- Plan y sus características.
- Variables demográficas y características del cliente.
- Equipo terminal y sus características.
- Cuotas pagadas de arriendo del equipo terminal.
- Minutos emitidos y recibidos.
- Mensajes emitidos y recibidos.
- Facturación del cliente.
- Cantidad de recargas y bolsas.
- Número de acciones de cobranza, suma de atrasos de pago y promedio de días de atraso los pagos.

- Tráfico de internet.

Para cada una de las variables numéricas, que representan algún tipo de tráfico o consumo, se crearon una serie de variables para agregar mayor información y predicción. Como ejemplo, se detallará las variables creadas para la facturación de los clientes.

- `prom_ingreso_1-3`: Promedio de la facturación desde el periodo t-1 (un periodo antes del up-selling o un periodo antes de Septiembre del 2011, según corresponda) hasta el periodo t-3.

- `prom_ingreso_4-6`: Promedio de la facturación de los clientes desde el periodo t-4 hasta el periodo t-6.

- `dif_ingresos_1-3_4-6`: Diferencia entre el promedio de los tres primeros meses de estudio con los últimos tres meses de estudio.

- `r_crec_ingresos_1-3_4-6`: Tasa de crecimiento en la facturación del cliente entre el promedio de los últimos tres meses con respecto a los primeros tres meses.

- `%_ingresos_plan`: Porcentaje del monto real facturado en los primeros tres meses con respecto al monto total del plan.

- `%_ingresos_plan_sinexceso`: Porcentaje del monto real facturado con respecto al monto total del plan, pero como máximo el 100%.

- `%_ingresos_plan_exceso`: Porcentaje sobre el 100% si es que el monto real facturado en los primeros tres meses fue mayor al monto total del plan.

- `categoría_exceso_ingresos`: Si el porcentaje de exceso de facturación con respecto al monto del plan está entre 0,1% y 20% corresponde a la categoría "1". Si el porcentaje se sitúa dentro del 21% y el 40% corresponde a la categoría "2". Si el porcentaje de exceso cae dentro del rango del 41% al 60% corresponde a la categoría "3". La categoría 4 corresponde al rango entre el 61% y el 80%, la categoría "5" entre el 81% y el 100%. Y por último la categoría "6" corresponde a los clientes que facturaron más del 100% en exceso de su plan.

- `%_ingresos_historico`: Porcentaje del monto real facturado en los primeros tres meses con respecto monto facturado promedio en los últimos tres meses de estudio.

- `%_ingresos_plan_sinexceso_historico`: Porcentaje del monto real facturado con respecto al monto facturado promedio en los últimos tres meses de estudio, pero como máximo el 100%.

- `%_ingresos_plan_exceso_historico`: Porcentaje sobre el 100% si es que el monto facturado en los primeros tres meses fue mayor al monto facturado promedio en los últimos tres meses de estudio.

- `categoría_exceso_ingresos_historico`: Si el porcentaje de exceso de facturación con respecto al monto facturado promedio en los últimos tres meses de estudio está entre 0,1% y 20% corresponde a la categoría "1". Si el porcentaje se sitúa dentro del 21% y el 40% corresponde a la categoría "2". Si el porcentaje de exceso cae dentro del

rango del 41% al 60% corresponde a la categoría “3”. La categoría “4” corresponde al rango entre el 61% y el 80%, la categoría “5” entre el 81% y el 100%. Y por último la categoría “6” corresponde a los clientes que facturaron más del 100% en exceso de su plan.

- **r_crec_ingresos_i**: Tasa de crecimiento mensual en la facturación de los seis meses en cuestión. Donde i representa el mes de la tasa de crecimiento, es decir, i toma los valores de 1 a 5.

- **r_crec_ingresos_prom**: Tasa de crecimiento promedio en la facturación del cliente.

La creación de variables para el resto de los atributos como mensajería, minutos, internet, recargas, bolsas y acciones de cobranza se realizó de manera análoga.

De esta manera se conformó el conjunto final de datos, en la figura 8 se muestra parte de la tabla final que contiene los datos objetivos para ser analizada y ser explotada por los modelos de minería de datos.

Figura 8: Conjunto de datos finales.

	movil character	rut character va	es_pt_(t) character v	plan_up-selling_(t-1) character varying(255)	forma_plan_(t-1) character varying	MM_(t-1) character v	familia_(t-1) character va	ilimitado_(t-1) character varying	CF_(t-1) character v	Min_(t-1) character va	id_usuario_es character varyin	genero character	estado_civil character va	nivel_estudio character varyi	actividad character va
1	62019942	122233960	0	290 CTA CONTROLADA	CC	0	NET	0	6990	114	-1	0	-1	-1	-1
2	62020018	95433715	0	710 Plan RED 440	SUS	0	NET	0	16490	440	-1	1	1	2	2
3	62020121	06156070K	0	733 CC PLANO 22500	CC	0	FLAT	0	22500	239	-1	1	-1	-1	6
4	62020627	162341774	0	185 CTA CONTROLADA	CC	0	FLAT	0	11000	65	-1	0	0	1	1
5	62020674	65378140	1	712 Plan RED 1100	SUS	0	NET	0	26290	1100	-1	1	-1	-1	6
6	62018051	14590210K	0	290 CTA CONTROLADA	CC	0	NET	0	6990	114	-1	1	1	1	2
7	62029688	147387813	0	185 CTA CONTROLADA	CC	0	FLAT	0	11000	65	-1	1	-1	-1	-1
8	62030024	142446685	0	185 CTA CONTROLADA	CC	0	FLAT	0	11000	65	-1	-1	-1	-1	-1
9	62030343	127611769	0	185 CTA CONTROLADA	CC	0	FLAT	0	11000	65	-1	0	0	3	2
10	62032800	217429781	0	291 CTA CONTROLADA	CC	0	NET	0	8990,45	169	-1	1	-1	-1	-1
11	62032928	166982324	1	290 CTA CONTROLADA	CC	0	NET	0	6990	114	-1	0	0	-1	0
12	62033323	93495195	0	290 CTA CONTROLADA	CC	0	NET	0	6990	114	-1	0	-1	-1	-1
13	62034736	94812399	0	290 CTA CONTROLADA	CC	0	NET	0	6990	114	-1	-1	-1	-1	-1
14	62021446	133439420	0	734 Familia Ilimitada	SUS	0	FLAT	1	30490	240	-1	0	0	3	2

Fuente: Elaboración Propia.

8.4 EXPLORACIÓN DE LOS DATOS

Recordando que los objetivos que se ha propuesto la compañía son masificar la experiencia multimedia y aumentar los ingresos promedio por usuario, es esencial traducir esto en un modelo de propensión de compra hacia los planes multimedia que ofrece la empresa. Los planes multimedia poseen mayor cargo fijo y mayor innovación tecnológica. Por lo tanto, son los planes que generan mayor valor para la compañía y del mismo modo mayor valor para los clientes.

En la figura 9 se presentan los doce Planes Multimedia (en adelante Planes Target), a los cuales se desea que los clientes realicen el Up-Selling.

Figura 9: Planes Multimedia.

PLAN TARGET	CARGO FIJO
1049 Multimedia CtaCda 15990	\$ 15.990
1253 Multimedia 85	\$ 18.480
1050 Multimedia 100	\$ 19.990
1205 Multimedia CtaCda 21.990	\$ 21.990
1051 Multimedia 140	\$ 22.490
1057 Multimedia Red 800	\$ 23.990
1052 Multimedia 200	\$ 25.990
1053 Multimedia 280	\$ 29.990
1058 Multimedia Red 1200	\$ 31.490
1054 Multimedia 400	\$ 39.990
1055 Multimedia 500	\$ 45.990
1056 Multimedia 800	\$ 59.990

Fuente: Elaboración Propia.

La participación, medida en frecuencia, de los Planes Target (PT) dentro del total de clientes corresponde a un 12,47%.

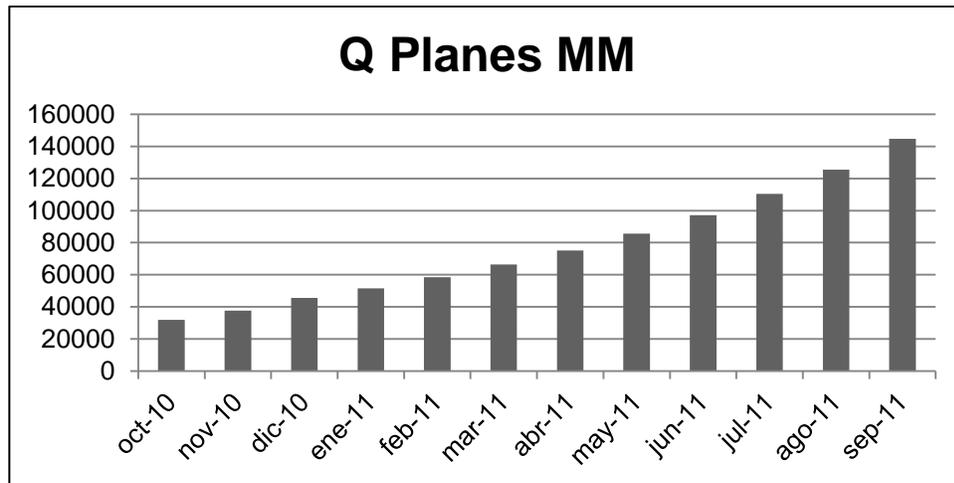
Figura 10: Participación en frecuencia de los Planes Multimedia.

PLAN TARGET	CARGO FIJO	FRECUENCIA	PORCENTAJE
1053 Multimedia 280	\$ 29.990	24242	2,09%
1058 Multimedia Red 1200	\$ 31.490	23854	2,05%
1057 Multimedia Red 800	\$ 23.990	22557	1,94%
1050 Multimedia 100	\$ 19.990	15272	1,32%
1054 Multimedia 400	\$ 39.990	14538	1,25%
1049 Multimedia CtaCda 15990	\$ 15.990	12536	1,08%
1055 Multimedia 500	\$ 45.990	11883	1,02%
1052 Multimedia 200	\$ 25.990	7228	0,62%
1056 Multimedia 800	\$ 59.990	7031	0,61%
1051 Multimedia 140	\$ 22.490	4323	0,37%
1205 Multimedia CtaCda 21.990	\$ 21.990	1292	0,11%
1253 Multimedia 85	\$ 18.480	0	0,00%
PARTICIPACION PT'S		144756	12,47%

Fuente: Elaboración Propia.

Es importante destacar que los Planes Multimedia fueron lanzados al mercado en Agosto del 2010. Durante el último año han tenido una tasa de crecimiento mensual cercana al 15%, que se ve reflaja en la figura 11. Esto indica que la oportunidad de mercado es gigante y que los desafíos por masificar la experiencia multimedia son alentadores.

Figura 11: Crecimiento en participación de Planes Multimedia.



Fuente: Elaboración Propia.

Con respecto a los ingresos de la compañía (facturación de los clientes), los PT tienen una mayor participación y esto se puede apreciar en la figura 12, donde “0” son los Planes No Target (PNT) y “1” son los PT.

Figura 12: Participación en ingresos de Planes Multimedia.

ES_PT	INGRESO	PORCENTAJE
0	\$ 1.887.930.063	81,30%
1	\$ 434.106.781	18,70%
Total general	\$ 2.322.036.844	100,00%

Fuente: Elaboración Propia.

La compañía posee un universo de 330 planes aproximadamente. Por ende, que 12 planes tengan el 18,7% de los ingresos habla de una gran fuente de ingreso.

8.4.1 TIPOS DE COMPORTAMIENTOS EN LOS PLANES

Dado que el objetivo es predecir el up-selling a los PT, es conveniente analizar el comportamiento de los clientes en la contratación de los planes durante la data histórica. Para ello se identificaron 8 tipos de comportamientos.

1. **Cambio entre PT:** Un cliente con PT se cambia a otro PT en el periodo siguiente.
2. **Mantuvo PT:** Un cliente con PT mantiene el plan PT en el periodo siguiente.

3. **Nuevo (o suspendido) a PT:** Un cliente nuevo (o suspendido) accede a un PT.
4. **Up-Selling:** Un cliente que posee un PNT se cambia a un PT en el periodo siguiente.
5. **Down-Selling:** Un cliente que posee un PT se cambia a un PNT en el periodo siguiente.
6. **Nuevo (o suspendido) a PNT:** Un cliente nuevo (o suspendido) accede a un PNT.
7. **Mantuvo PNT:** Un cliente con PNT mantiene el plan PNT en el periodo siguiente.
8. **Cambio entre PNT:** Un cliente con PNT se cambia a otro PNT en el periodo siguiente.

Cada uno de estos comportamientos, que se aprecian en la figura 13, se etiquetó de la siguiente manera para un mejor manejo en los análisis.

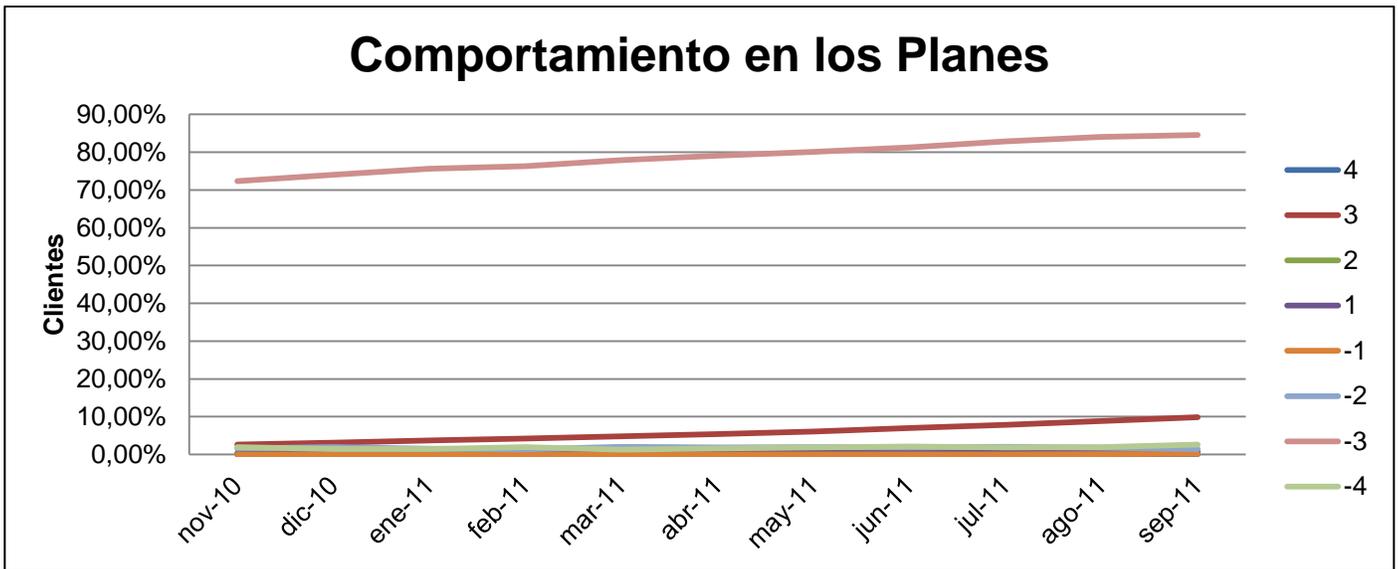
Figura 13: Tipos de Comportamientos.

Tipos de Comportamientos	
4	Cambio entre PT
3	Mantuvo PT
2	Nuevo (o suspendido) a PT
1	Up-selling
-1	Down-selling
-2	Nuevo (o suspendido) a PNT
-3	Mantuvo PNT
-4	Cambio entre PNT

Fuente: Elaboración Propia.

El 87,53% de los clientes posee un PNT. Observando su historia, se puede identificar un comportamiento de status quo con los planes PNT (-3). Esto mismo hace que el resto de las curvas de comportamiento sean poco claras o bastante pequeñas en participación, como se puede apreciar en la figura 14.

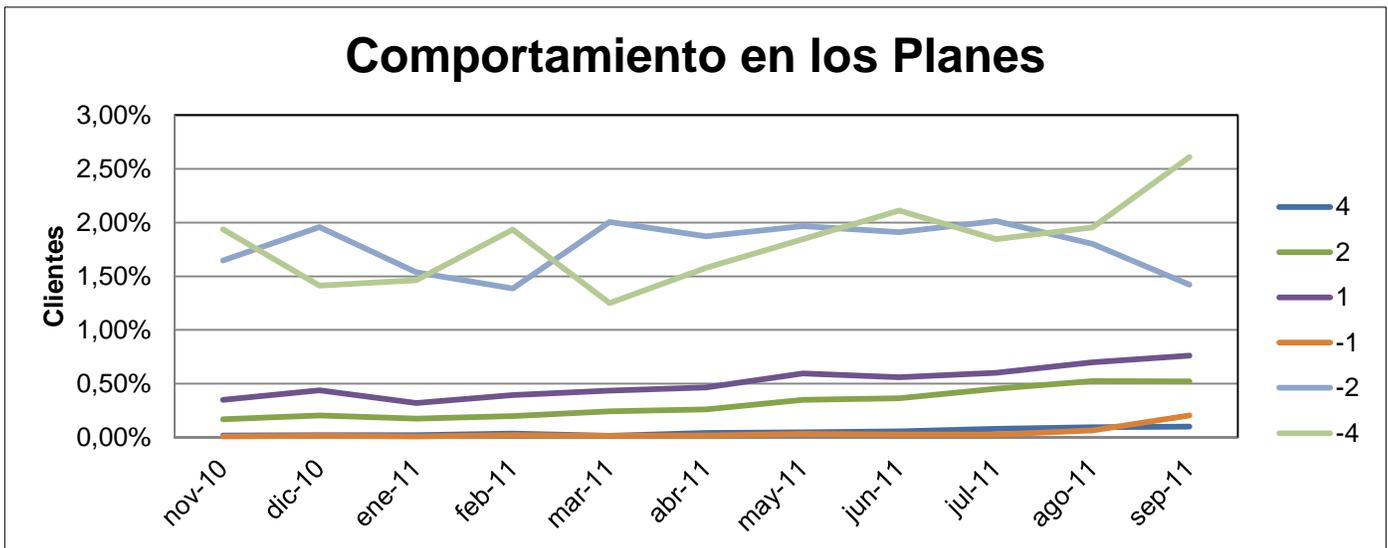
Figura 14: Curvas de comportamiento de los clientes.



Fuente: Elaboración Propia.

Lo que se desea encontrar son patrones que caractericen a los clientes que han realizado up-selling. Por lo tanto se realizara un zoom al gráfico anterior, para identificar con mayor claridad la participación de los up-sellers (1). En la figura 15, se realiza un zoom con respecto a la figura 14.

Figura 15: Zoom al comportamiento de los clientes.



Fuente: Elaboración Propia.

Como la mayoría de los clientes posee PNT, prevalece esta situación al momento de analizar los cambios de planes, y se aprecia una mayor participación de

clientes nuevos a PNT (-2) y de cambios entre PNT (-4). Pero dejando de lado el sesgo incluido en el análisis se puede observar un crecimiento paulatino en los up-sellers, los cuales son los clientes que nos interesan y analizaremos en los puntos siguientes.

8.4.2 TIPOS DE COMPORTAMIENTOS EN LOS PLANES DE UP-SELLERS

De aquí, en adelante solo se incluirá a los clientes que dentro del año de estudio, en algún momento del periodo, realizaron un up-selling. En la figura 16 podemos apreciar la misma categorización mostrada con anterioridad, pero en este caso solo está orientada para los clientes que han realizado up-selling.

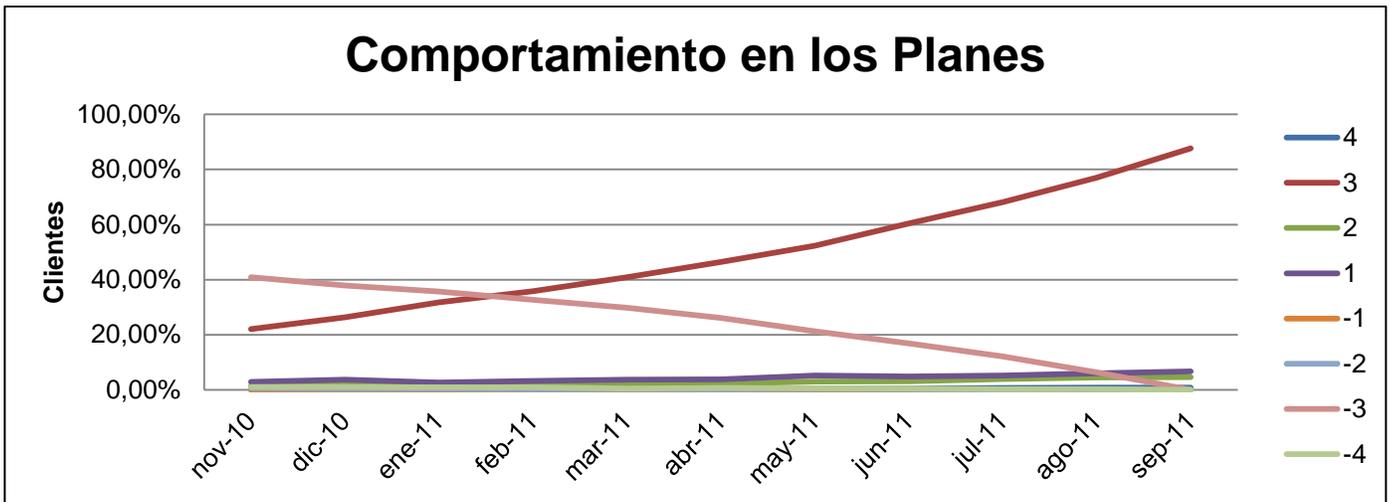
Figura 16: Tipos de comportamientos enfocados a clientes up-seller.

Tipos de Comportamientos	
4	Cambio entre PT
3	Mantuvo PT
2	Nuevo (o suspendido) a PT
1	Up-selling
-1	Down-selling
-2	Nuevo (o suspendido) a PNT
-3	Mantuvo PNT
-4	Cambio entre PNT

Fuente: Elaboración Propia.

En la figura 17, se observa una clara tendencia, en que los clientes que accedieron a un PT, antes del periodo de estudio (3), se han mantenido en ellos y de manera contraria los clientes que tenían un PNT (-3) se han ido cambiando a PT en la misma proporción.

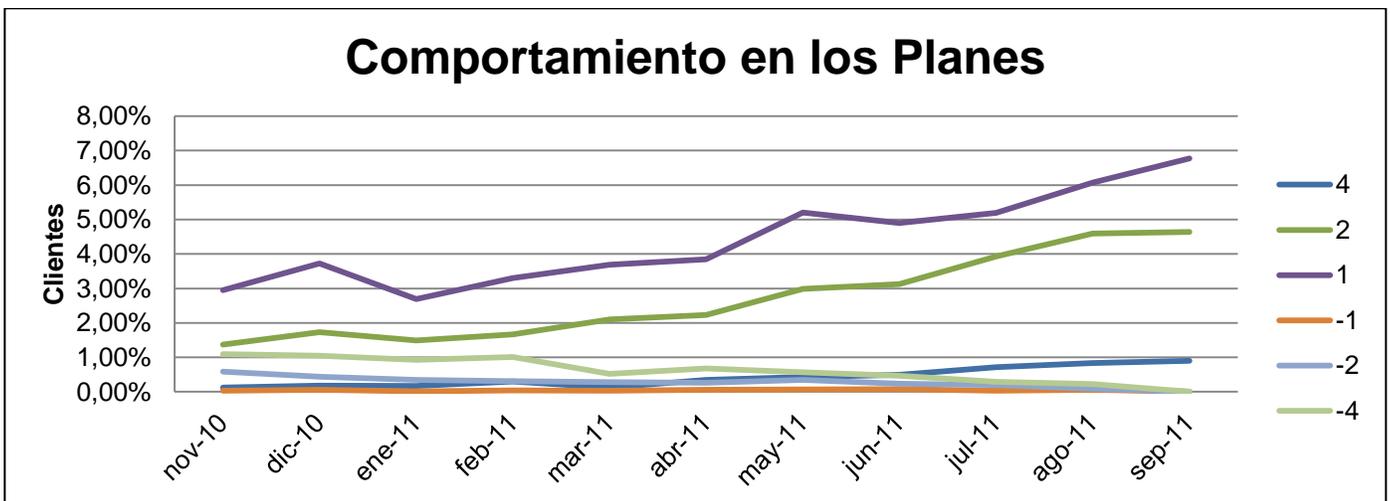
Figura 17: Curvas de comportamiento para clientes up-seller.



Fuente: Elaboración Propia.

El comportamiento que interesa es el del Up-Selling (1) por lo que se hará un zoom para poder notar el resto de las curvas. Se puede apreciar, en la figura 18, que los clientes que realizan Up-Selling (1) ha ido aumentando con el tiempo de manera bastante explosiva. Del mismo modo, los clientes nuevos (o suspendidos) en mayor proporción acceden a PT (2). Y para el análisis es grato encontrar que los comportamientos relacionados con PNT están decreciendo y su participación es bastante inferior a los comportamientos relacionados con PT.

Figura 18: Zoom al comportamiento de los clientes up-seller.



Fuente: Elaboración Propia.

Basándose en lo anterior, es de vital importancia comprender que hace que los clientes tomen la decisión de cambiar desde un PNT a un PT y realizar un Up- Selling. Es por esto que a continuación se mostraran las variables relevantes que caracterizan a estos clientes.

8.4.3 ANÁLISIS DEMOGRÁFICO Y GEOGRÁFICO DE LOS UP-SELLERS

En esta instancia se analizaron las relaciones entre las distintas variables demográficas y geográficas con la variable dependiente. La variable dependiente es una variable dummy, la cual es “0” si el cliente no ha realizado up-selling y “1” si es que realizó. El análisis de relación entre la variable dependiente y las variables explicativas se hace empleando tablas de contingencia, de manera descriptiva, sin involucrar las pruebas estadísticas de asociación e independencia. Es decir, un análisis basado en los porcentajes y recuentos.

Para el atributo “Género”, se puede analizar la participación dentro de los PNT (0) y de los PT (1). Cada columna equivale a una categorización del género, siendo -1 No Informado, 0 Femenino y 1 Masculino.

Figura 19: Distribución del género.

GENERO	-1	0	1	Total
0	9,14%	41,83%	49,03%	100,00%
1	9,70%	32,77%	57,53%	100,00%

Fuente: Elaboración Propia.

Se puede apreciar que el 57,53% de los clientes que poseen PT son Hombres y que solamente el 32,77% son Mujeres.

Analizando la tasa de captación de los planes dentro del género, podemos observar la siguiente tabla.

Figura 20: Tasa de captación del género.

GENERO	0	1
0	89,7%	85,3%
1	10,3%	14,7%
Total	100%	100%

Fuente: Elaboración Propia.

Los PT poseen el 10,34% de la Mujeres que son clientes de la compañía y el 14,73% de los clientes que son hombres. Se puede inferir que los hombres han estado más inclinados a realizar Up-Selling.

Si no se consideran los datos no informados, el resultado es el siguiente.

Figura 21: Distribución del género sin valores perdidos.

GENERO			
	Mujer	Hombre	
UP-SELLER	0	1	TOTAL
0	46,09%	53,91%	100,00%
1	39,96%	60,04%	100,00%
TOTAL	45,85%	54,15%	100,00%

Fuente: Elaboración Propia.

Con este análisis, es más claro identificar que la proporción de hombres que realizan up-selling es mucha mayor que la proporción de mujeres que lo realiza. En este sentido, se puede afirmar que los hombres son los que están siendo líderes en la contratación de Planes Multimedia.

Para el atributo “Edad”, también se analizó la participación dentro de los PNT (0) y de los PT (1). Cada columna equivale a un rango de edad, siendo NI no informado.

Figura 22: Distribución de la edad.

EDAD	NI	7-14	15-20	21-30	31-40	41-50	51-60	61-70	71-110	Total
0	10,98%	0,54%	3,06%	18,20%	23,22%	19,75%	14,61%	6,94%	2,69%	100,00%
1	11,46%	0,20%	2,29%	32,68%	31,16%	12,14%	6,95%	2,37%	0,74%	100,00%

Fuente: Elaboración Propia.

La mayor cantidad de PT se encuentra entre los 21 y los 40 años. En cambio los PNT se encuentran repartidos desde los 21 hasta los 60 años. Este resultado es bastante intuitivo, ya que los PT son planes de tecnología avanzada, donde los clientes de mayor edad son más reacios a contratarlos.

Si ahora se analiza la tasa de captación según rango de edad para los PT, se puede apreciar el mismo fenómeno comentado con anterioridad.

Figura 23: Tasa de captación de la edad.

EDAD	7-14	15-20	21-30	31-40	41-50	51-60	61-70	71-110
0	94,79%	90,09%	79,10%	83,51%	91,70%	93,45%	95,21%	96,08%
1	5,21%	9,91%	20,90%	16,49%	8,30%	6,55%	4,79%	3,92%
Total	100%	100%	100%	100%	100%	100%	100%	100%

Fuente: Elaboración Propia.

Los PT poseen el 20,90% de los clientes entre los 21 y los 30 años de la compañía y el resto de los rangos de edad posee una tasa de captación muy baja. Se podría inferir que no es el producto para ese tipo de clientes o que se requiere una profundización de los datos para entender cuáles estarían más interesados a contratar.

Si no se consideran los datos no informados, el resultado es el siguiente.

Figura 24: Distribución de la edad sin valores perdidos.

UP-SELLER	EDAD									TOTAL
	0-6	7-14	15-20	21-30	31-40	41-50	51-60	61-70	71-110	
0	0,03%	0,59%	3,28%	19,78%	26,06%	22,38%	16,77%	7,99%	3,11%	100,00%
1	0,00%	0,16%	2,19%	34,26%	36,38%	14,53%	8,69%	3,07%	0,72%	100,00%
TOTAL	0,03%	0,57%	3,24%	20,35%	26,47%	22,07%	16,45%	7,80%	3,02%	100,00%

Fuente: Elaboración Propia.

Es claro notar que la mayor cantidad de up-seller tiene entre 21 y 40 años. El Plan Multimedia, entonces, apunta a personas con mayor conocimiento tecnológico, que coinciden con personas más jóvenes. En cambio, para los clientes que aún no realizan un up-selling, la proporción de edad es más distribuida, centrándose entre los 21 y los 60 años.

Para el atributo "GSE", que corresponde al sector económico, se analizó la participación de cada uno de los sectores económicos dentro de los PNT (0) y los PT (1).

Figura 25: Distribución del sector económico.

GSE	ABC1	C2	C3	D	E	NI	Total
0	14,98%	16,65%	17,14%	9,78%	0,51%	40,94%	100,00%
1	21,98%	15,58%	11,09%	5,23%	0,17%	45,95%	100,00%

Fuente: Elaboración Propia.

La mayor cantidad de PT lo posee el sector ABC1, superando al resto de los sectores por una gran diferencia. En cambio, para los PNT, la participación está bastante repartida entre los segmentos ABC1, C2 y C3. Este resultado también resulta intuitivo, ya que los precios de los planes multimedia (PT) son más elevados en comparación con los PNT. Además, el equipo también juega un rol importante dentro de la decisión, puesto que no solo el precio del plan influye, sino que también el precio de Smartphone.

Figura 26: Tasa de captación del sector económico.

GSE	ABC1	C2	C3	D	E
0	82,24%	87,89%	91,30%	92,70%	95,27%

1	17,76%	12,11%	8,70%	7,30%	4,73%
Total	100,00%	100,00%	100,00%	100,00%	100,00%

Fuente: Elaboración Propia.

Los PT poseen el 17,76% de los clientes ABC1 de la compañía y de los sectores económicos posee una tasa de captación muy baja. Se podría inferir que no es el producto para ese tipo de clientes o que se requiere una profundización de los datos para entender cuáles estarían más interesados a contratar.

Si no se consideran los datos no informados, el resultado es el siguiente.

Figura 27: Distribución del sector económico sin valores perdidos.

GSE						
UP-SELLER	ABC1	C2	C3	D	E	TOTAL
0	25%	28%	29%	17%	1%	100%
1	40%	30%	21%	9%	0%	100%
TOTAL	26%	28%	29%	16%	1%	100%

Fuente: Elaboración Propia.

Claramente existe una tendencia a que los up-seller pertenezcan a estratos socioeconómicos más altos, encontrando su máximo en el segmento ABC1. Para los clientes que no realizan up-selling este más distribuido, encontrando su máximo en el segmento C3.

Para poder determinar si existe una relación entre dos variables no basta con los porcentajes o los recuentos, la relación debe ser comprobada mediante una prueba estadística. Para ello se ocupara el Test Chi Cuadrado con tablas de contingencia. Es necesario resaltar que esta prueba nos indica si existe, o no, una relación entre las variables, pero no indica el grado o el tipo de relación.

Figura 28: Pruebas Chi-cuadrado.

Pruebas de chi-cuadrado		
Variable	Test	Sig. asintótica (bilateral)
Género	Chi-cuadrado de Pearson	0,00
Edad	Chi-cuadrado de Pearson	0,00
GSE	Chi-cuadrado de Pearson	0,00

Fuente: Elaboración Propia.

Como los valores de la Significancia Asintótica son menores a 0,05 se rechaza la hipótesis del test. Es decir, las variables expuestas están relacionadas con la variable dummy “es_pt” (1 si lo posee y 0 si no).

El mismo procedimiento se realizó para el resto de las variables como estado civil, nivel de estudio, actividad, profesión, posee hijos, tipo de usuario y región. Pero para estas variables la cantidad de datos no era representativa, es decir estaba poco poblada cada una de las variables, por lo tanto el resultado no entrega mucha información y calidad para un análisis exacto.

8.4.4 ANÁLISIS DE FACTURACIÓN DE UP-SELLERS

Para seguir con el análisis de los clientes que realiza up-selling, se estudió el comportamiento en la facturación durante el año de data. Este análisis contempla la facturación por los servicios contratados, las bolsas contratadas en los distintos servicios y las recargas realizadas mensualmente. La manera en que se realizó este análisis, se basa en la localización temporal del periodo t donde los clientes realizaron el up-selling. Este periodo t será el punto de inflexión a partir del cual se analizaran las diferencias entre el comportamiento pasado y el futuro. Resulta sesgado analizar todos los clientes que realizaron up-selling de manera homogénea, ya que cada uno de estos posee planes diferentes, que se distinguen en cargo fijo, minutos, mensajes y otros. Por lo tanto, la manera de hacer comparables a todos los up-sellers es definir una facturación base, la cual se calcula como el promedio de facturación del periodo t-4 hasta completar el año. Es decir, el promedio de facturación de 4 meses antes del up-selling hasta que el cliente posea datos. Luego de calcular la base de facturación se calcula el porcentaje de facturación compara con la base del resto de los meses.

El figura 29 ilustra el comportamiento en el gasto de los clientes, en la vecindad del up-selling.

Figura 29: Curva de ingresos para la compañía (up-sellers).

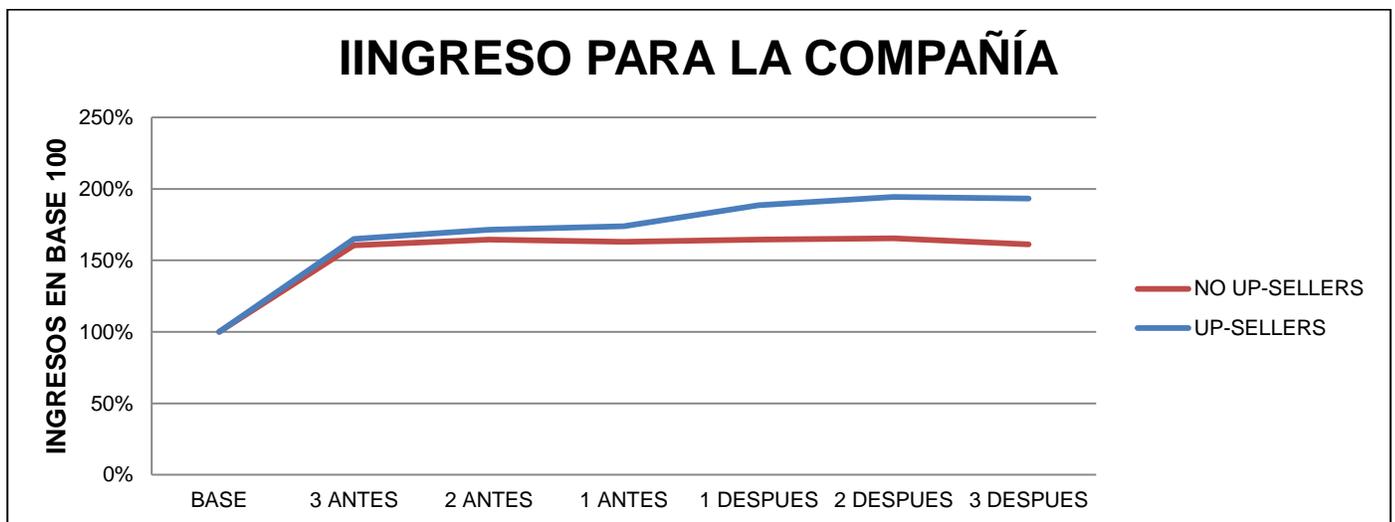


Fuente: Elaboración Propia.

Es posible apreciar, que los clientes que realizaron up-selling venían gastando sobre su promedio en los últimos periodos. Y en el momento de realizar el up-selling la facturación de los clientes siguió aumentando. Aquí se podrían intuir muchas cosas, como por ejemplo que los clientes que van gastando más de lo normal deciden cambiar a un mejor plan para que el costo marginal de su consumo sea menor. Por otro lado se podría pensar que la tendencia del mercado hace que los clientes vayan gastando más con el transcurso del tiempo. Sin embargo, no se tiene mucha información para dar una respuesta correcta, lo que sí se puede concluir es que en la vecindad del up-selling existe un crecimiento en la facturación.

Además de estudiar a los clientes que realizaron up-selling es necesario analizar el comportamiento de quienes aún no lo han hecho. Pues, esta tendencia puede ser una condición de mercado y los dos tipos de clientes pueden tener las mismas características de facturación. Es por esto, que a continuación se muestra la diferencia de las curvas de facturación para estos comportamientos.

Figura 30: Curva de ingresos para la compañía (up-sellers vs no up-sellers).



Fuente: Elaboración Propia.

La interpretación que se puede extraer es favorable para el estudio, ya que los clientes que realizan up-selling presentan un aumento superior (cercano al 40%) con respecto a los que no realizan up-selling. Por lo tanto se puede extraer una característica diferenciadora de los up-sellers, su comportamiento en la facturación es creciente y superior al de lo que no realizan up-selling, después de haber realizado un cambio de plan a un PT. De esta manera la variable facturación podría ser relevante dentro del estudio.

8.4.5 ANÁLISIS DEL TRÁFICO DE MINUTOS DE LOS UP-SELLERS

Para analizar el comportamiento del tráfico de minutos de los clientes que realizaron up-selling, se siguió el mismo enfoque que para el caso anterior. El siguiente gráfico muestra el tráfico de minutos emitidos en el periodo de análisis, para los up-sellers en la vecindad del punto de inflexión.

Figura 31: Curva de tráfico de minutos para la compañía (up-sellers).

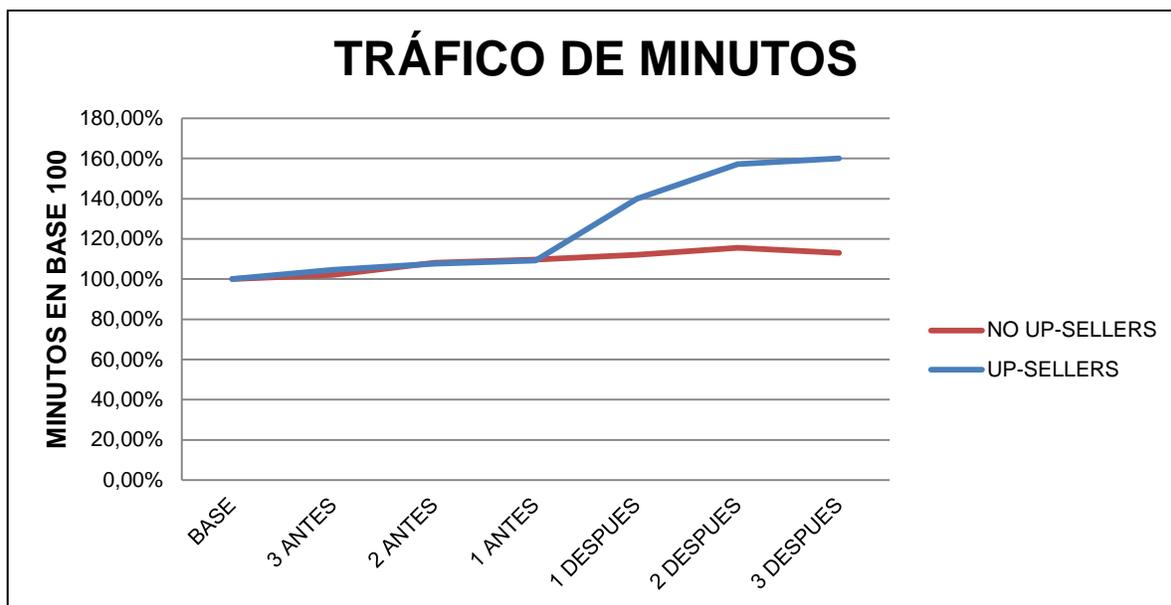


Fuente: Elaboración Propia.

Se puede observar que antes de realizar el up-selling, el tráfico de minutos crecía lentamente a una tasa constante. Pero luego de haber realizado el cambio de plan, el tráfico de minutos creció rápidamente en los siguientes dos periodos hasta mantenerse constante por sobre un 60% de lo traficado en el pasado.

Es necesario comparar los clientes que realizaron up-selling con los que no, para así poder inferir y respaldar ciertos patrones diferenciadores entre los dos tipos de clientes. Para esto, se muestra el siguiente gráfico que compara el tráfico de estos dos comportamientos.

Figura 32: Curva de tráfico de minutos para la compañía (up-sellers vs no up-sellers).



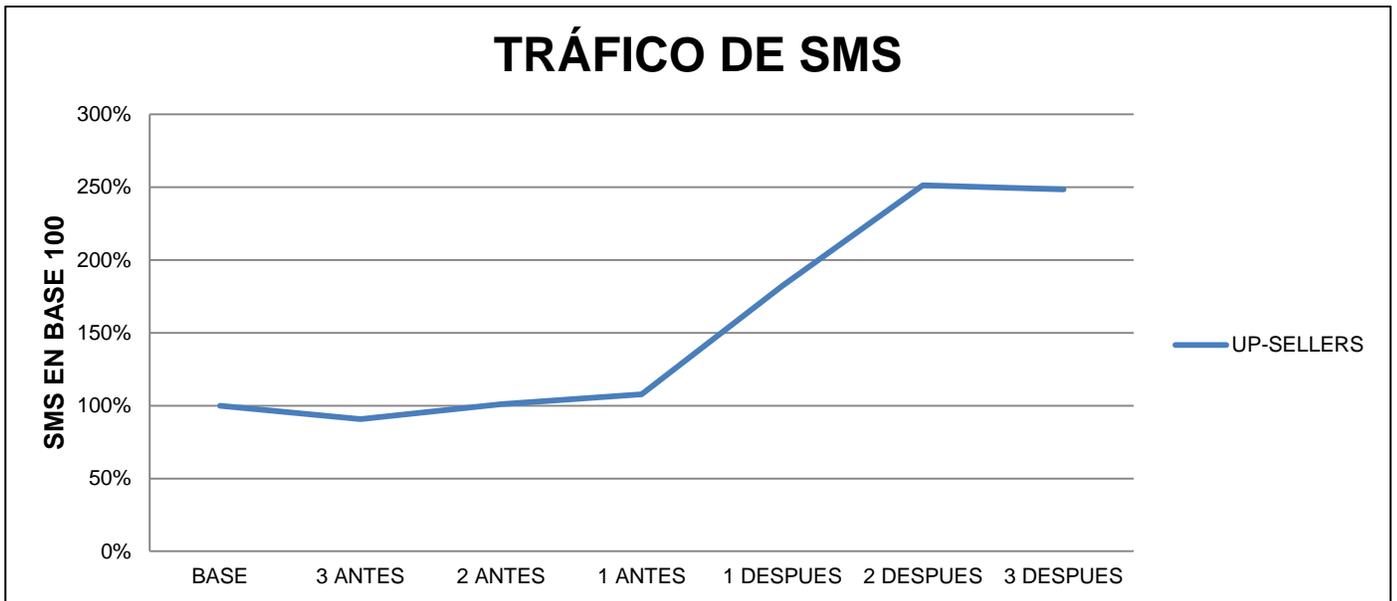
Fuente: Elaboración Propia.

De la figura 32 es posible interpretar que no existe gran diferencia entre los dos tipos de clientes, ya que la curva es muy similar por no decir igual hasta el momento en que los up-sellers realizan el up-selling. Luego de este punto, existe una marcada diferencia en las curvas, puesto que para los up-seller el crecimiento en tráfico de minutos es muy superior al de los no up-sellers. Pues los que no realizan up-selling siguen creciendo en tráfico pero a una tasa bastante menor que los up-sellers. Por lo tanto se puede identificar que todos los clientes mantienen un comportamiento similar en el tráfico de minutos, el cual es ir creciendo paulatinamente a una tasa constante, pero la diferencia se encuentra en los que realizaron up-selling y los que no, ya que el tráfico de up-sellers es muy superior al resto de los clientes.

8.4.6 ANÁLISIS DEL TRÁFICO DE MENSAJES DE LOS UP-SELLERS

El tráfico de mensajes se analizó de la misma manera que los estudios anteriores. De esta manera, lo primero fue analizar el tráfico de mensajes de los clientes que realizaron up-selling. En el siguiente gráfico se muestra cual es la tendencia que sigue el comportamiento de estos clientes.

Figura 33: Curva de tráfico de mensajes para la compañía (up-sellers).

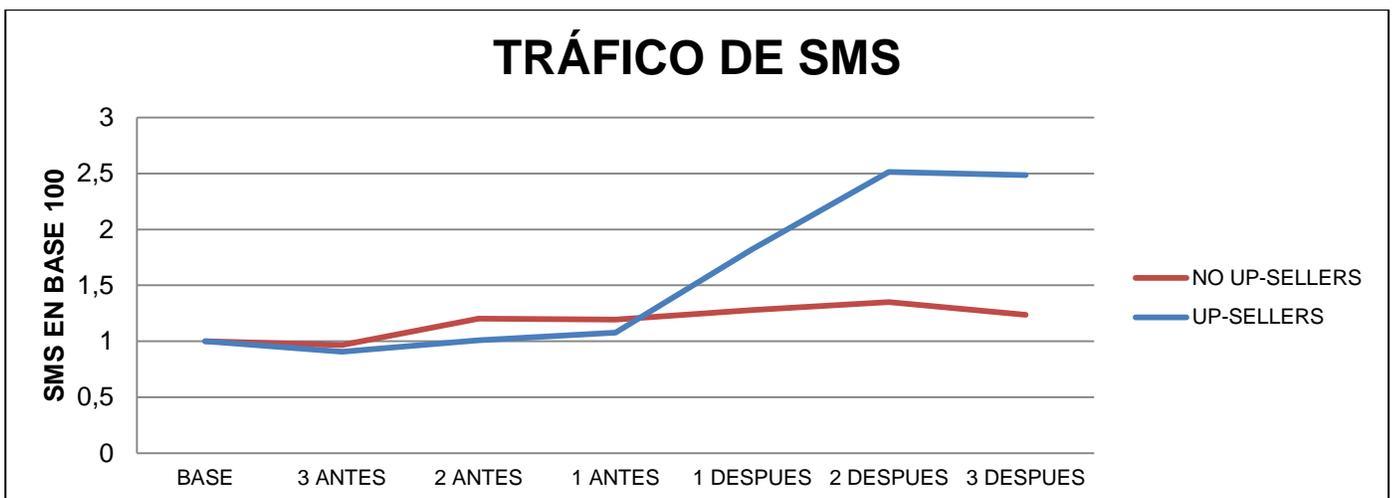


Fuente: Elaboración Propia.

Es posible apreciar que el tráfico se mantiene, relativamente, constante los periodos anteriores al up-selling, pero al momento de cambiar de plan el tráfico aumenta considerablemente en los siguientes dos periodos hasta mantenerse constante a través del tiempo.

Al contra restar los dos comportamientos, es decir, los clientes up-sellers y los no up-seller se obtiene el siguiente gráfico.

Figura 34: Curva de tráfico de mensajes para la compañía (up-sellers vs no up-seller).



Fuente: Elaboración Propia.

Se puede observar, en la figura 34 que el comportamiento para los dos tipos de clientes es similar hasta el momento del up-selling, pero la gran diferencia se vuelve a encontrar después de la realización del up-selling. Por lo tanto, no es posible distinguir una clara diferencia en el comportamiento anterior al up-selling para poder inferir quienes son los que cambiara a un PT o no. Pero si es posible percatar que el cambiar a un plan multimedia el tráfico de mensajes aumenta considerablemente en relación a los que no lo han hecho. Posiblemente ese aumento de mensajes se debe a que el plan incluye mensajes como parte de la oferta.

8.4.7 ANÁLISIS DE EQUIPOS CELULAR DE LOS UP-SELLERS

Lo que se busca analizar con este estudio es poder observar cual es el patrón de comportamiento con los equipos celulares de los clientes que realizan up-selling, es decir, identificar cuando los clientes cambian su equipo a un Smartphone. La razón de estudiar este comportamiento se traduce en la importancia que tiene el equipo celular al momento de tomar la decisión de contratar un plan multimedia.

Para analizar esta variable se identifican 6 tipos de comportamientos:

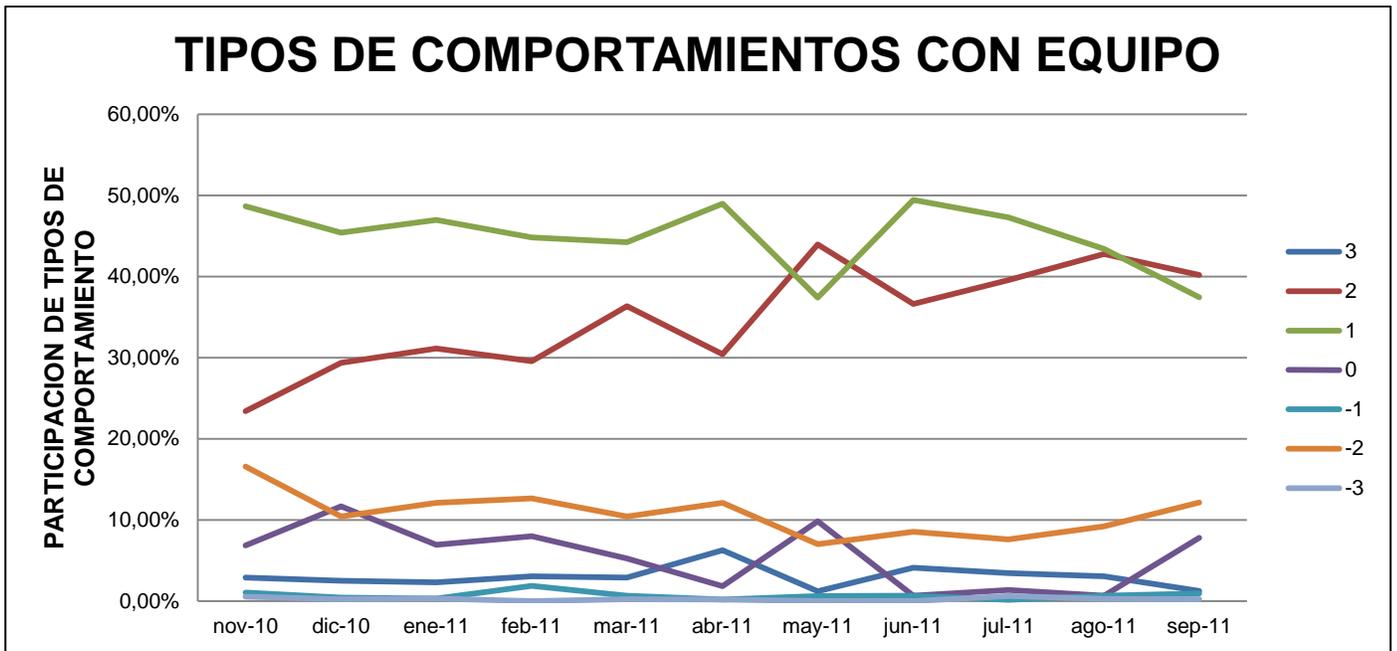
Figura 35: Tipos de comportamientos con los equipos terminales.

TIPOS DE COMPORTAMIENTOS	
3	Posee Smart para el Up-Selling (no se sabe si fue en conjunto o antes)
2	Posee Smart desde antes del Up-Selling
1	Adquirió Smart junto con Up-Selling
0	No se puede inferir
-1	Bajo de Tecnología junto con Up-Selling
-2	No posee Smart
-3	No posee Smart para el Up-Selling (no se sabe si fue en conjunto o antes)

Fuente: Elaboración Propia.

Para los clientes que han realizado up-selling, durante el año de estudio, se analizó su comportamiento a través del tiempo, obteniendo el siguiente gráfico como resultado.

Figura 36: Curvas de comportamiento equipos terminales.

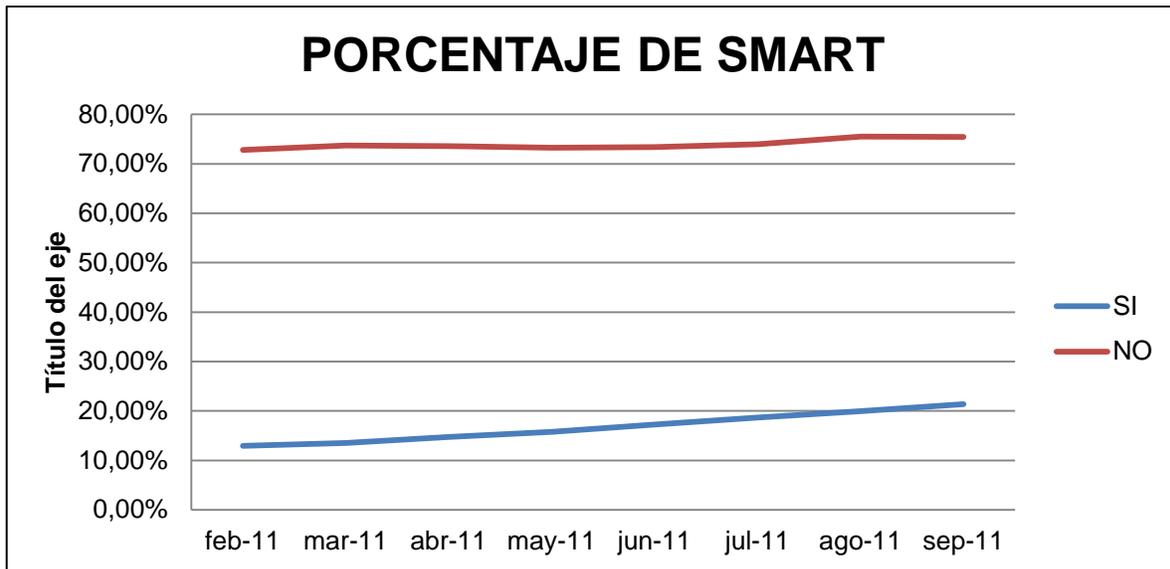


Fuente: Elaboración Propia.

Es posible observar que existen dos tendencias predominantes. La primera es comprar el Smartphone junto a la contratación de un PT y la segunda es poseer un Smartphone antes de realizar la contratación de un PT. Dentro del año de estudio es posible observar que en Noviembre del 2010 los clientes, en su mayoría, contrataban un PT en conjunto con la compra de un Smartphone, pero este patrón fue cambiando a través del tiempo hasta que en Septiembre del 2011 la mayoría de los clientes ya habían realizado el cambio de celular a un Smartphone, antes de contratar un PT. El argumento más intuitivo para responder a este patrón de comportamiento, es que los Smartphone fueron equipos poco accesibles en su comienzo, por el elevado costo que los caracterizaba. Pero durante el transcurso del año, la gama de productos Smartphone aumento, y las posibilidades de tener uno se incrementaron en conjunto con los descuentos asociados al equipo por acceder a un Plan Multimedia. Del mismo modo, la red de contactos de todas las personas aumento en equipo terminales más tecnológicos, lo que genera una presión extra para que la persona siga a los líderes de la innovación.

Por otra parte, es importante identificar el comportamiento con los equipos terminales de los clientes que no han realizado up-selling. El siguiente gráfico muestra el porcentaje de clientes que tiene samrtphone y el de los que no tienen dentro del total de clientes que aún no contratan un PT.

Figura 37: Porcentaje de Smart para clientes sin Plan Multimedia.



Fuente: Elaboración Propia.

El resultado es bastante intuitivo, debido a que el porcentaje de clientes que poseen Smartphone dado que no tienen planes multimedia es muy bajo. Pero esta tendencia ha ido cambiando a través del tiempo y el porcentaje aumento aproximadamente hasta el 21%, lo que no es una cifra baja.

8.4.8 RESULTADOS EN LA COMPRESIÓN DE LOS DATOS

En conclusión y abarcando las variables que no fueron expuestas anteriormente. Los clientes que han realizado up-selling en su mayoría son hombres, entre 21 y 40 años, de sector económico ABC1 o C2, solteros, con estudios Universitarios, de actividad profesional Ingenieros, Minería, Salud, Docencia y Administrativos, con uno a cuatro años de antigüedad en la compañía.

Del mismo modo, se puede decir que son clientes que han ido aumentando su facturación en los últimos periodos hasta traducir este comportamiento en un salto drástico en su facturación al momento de realizar el up-selling. Para el tráfico de minutos y el tráfico de mensajes existe una similitud con lo anterior. Su tráfico aumenta con el transcurso del tiempo, sin una gran diferencia con los clientes que no han realizado up-selling, pero al momento de realizar el cambio de plan su tráfico aumenta considerablemente.

Analizando la compra de equipos Smartphone, se puede rescatar que los clientes que contratan un PT en su mayoría compran un Smartphone, ya sea juntó con la

contratación o un par de periodos antes. Esta tendencia ha ido cambiando durante el tiempo, y últimamente el cambio se está realizando antes de la contratación del PT.

8.5 MINERÍA DE DATOS

Para la aplicación de los modelos, se generaron dos muestras de datos. En primer lugar se consideró a todos los clientes que tenían contratado un plan Suscripción (SUS). Estos planes se caracterizan por ser ilimitados en consumo de los servicios entregados por la compañía, es decir, el cliente que posee un plan Suscripción puede exceder el consumo de minutos, mensajes y descarga de datos que caracterizan a su plan.

Como segunda muestra se consideró a todos los clientes que tenían contratado un plan Cuenta Controlada (CC). Estos planes se caracterizan por tener un costo fijo asociado a minutos, mensajes y descarga de datos, pero con la salvedad de no poder exceder en cada uno de los límites de consumo establecidos por el plan.

La justificación de separar estos dos tipos de cliente, se basa en que tienen características y comportamientos distintos. Pues un cliente con un plan Suscripción posee un consumo más elevado de los servicios y si es que no lo realiza en la práctica está dispuesto a pagar más por tener la posibilidad de realizarlo. En el caso de los clientes Cuenta Controlada, su comportamiento es bastante restrictivo y su uso está orientado al autocontrol.

La aplicación de las técnicas de minería de datos se realizó bajo la herramienta RapidMiner 5.2. Esta herramienta se caracteriza por modelar procesos de Minería de Datos de manera simplificada, maneja elevadas cantidades de datos, es libre, excelente visualización tanto en los datos como en los modelos, trabaja en multi plataformas y posee una representación interna de los procesos de análisis de datos en ficheros XML.

8.5.1 APLICACIÓN DE ÁRBOL DE DECISIÓN A CLIENTES SUS

El conjunto de datos objetivos de los clientes Suscripción tiene la característica que un 13,98% del total ha realizado up-selling. Por ende, el restante 86,02% corresponden a clientes que se mantuvieron en planes distintos a los Multimedia. Lo que implica que el conjunto de datos final está totalmente desbalanceado.

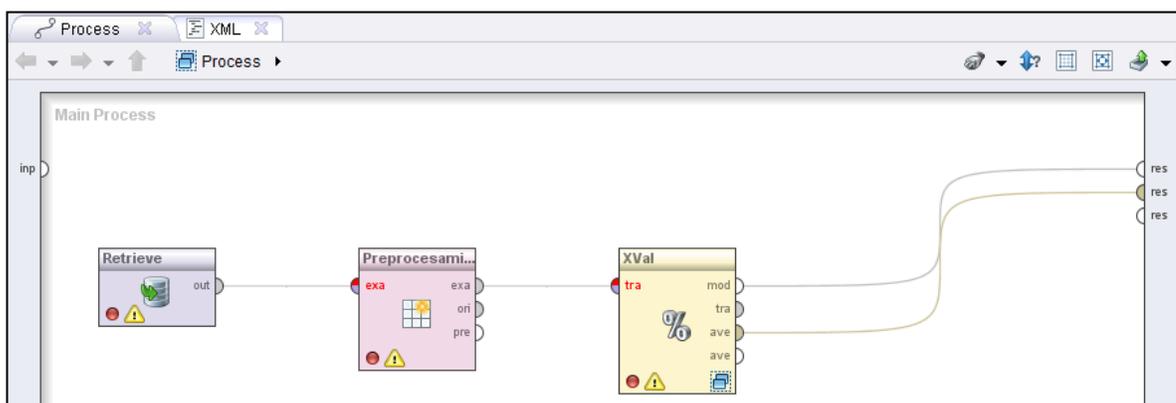
Para solucionar este problema se dividió la muestra en dos conjuntos de datos. La primera es la Muestra de Entrenamiento, la cual posee los datos con los que se entrenarán los modelos. La segunda es la Muestra de Prueba, donde se selecciona el mejor de los modelos entrenados y se analizan medidas de error en la predicción.

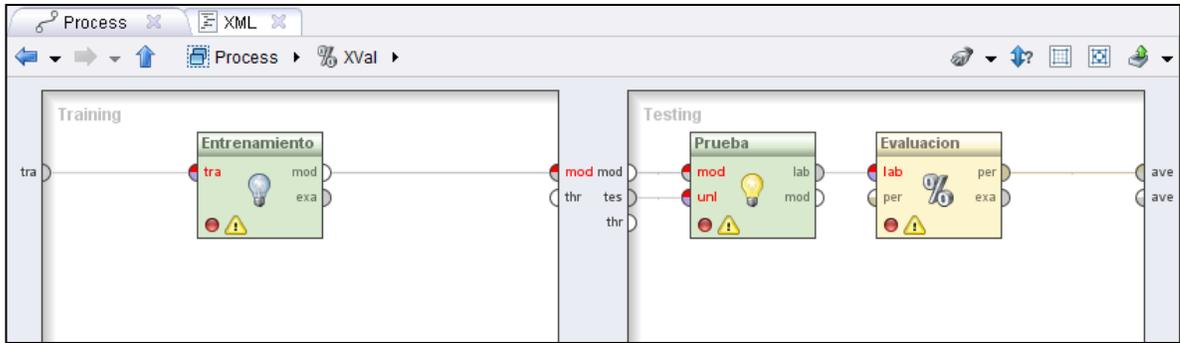
Esta separación se debe a que en la muestra de entrenamiento, los modelos se pueden ajustar a los datos, por lo tanto resulta necesario observar medidas de error con datos distintos a los utilizados en el entrenamiento. La muestra de entrenamiento tiene como característica la proporcionalidad de las cantidades, de tal manera de generar un muestreo estratificado compuesto por un 50% de clientes que realizaron up-selling y 50% de clientes que aún no lo han realizado. La separación entre la muestra de entrenamiento y la muestra de prueba se realiza usando la proporción 70-30. Esto quiere decir que la muestra de entrenamiento tendrá el 70% del universo de datos y la muestra de prueba contendrá el 30% restante. Ahora bien, la proporción de clientes en la muestra de prueba vuelven a ser la proporción real, de decir, 13,98% para los up-sellers y 86,02% para los no up-sellers.

El árbol de predicción se creó utilizando todos los atributos o variables que caracterizan cada uno de los registros en estudio. La variable independiente de este modelo será la variables “**es_pt_(t)**” que corresponde a “1” si es up-seller y “0” si no lo es. Por lo tanto, la variable a predecir es de tipo categórico.

El proceso dentro de la herramienta de RapidMiner, que se puede apreciar en la figura 38, parte con el input de los datos, el procesamiento de limpieza de valores perdidos y la validación cruzada que aplica el árbol de decisión.

Figura 38: Proceso de modelados RapidMiner.





Fuente: Elaboración Propia.

Se aplicaron distintos tipo de árboles, con distintos tipos de parámetros, hasta llegar a un árbol donde se obtuvieran los mejores resultados.

Se corroboró a través del método de “Bootstrapping” la distribución del error en el muestreo de entrenamiento. Esto significa que se crearon 10 muestras de entrenamiento aleatorias para aproximar la varianza de la predicción de los modelos. Entregando una varianza de 0,09%, lo que indica que el modelo es confiable para distintas muestras de datos.

Al aplicar el árbol de decisión a la muestra de prueba, la herramienta de minería de datos permite elegir el criterio con el cual se seleccionaran los atributos en casa uno de los nodos. Probando con todos los criterios posibles, el que mejor resultados entrego fue el criterio de “Ganancia de Información”, que permite pre y post poda, y además no tendrá problemas en penalizar atributos con mayor cantidad de datos, ya que con el procesamiento de limpieza todas las variables contienen la misma cantidad de datos.

Se utilizó validación cruzada, es decir, el árbol generó subconjuntos de datos de la muestra total para estimar la predicción del árbol construido sobre el conjunto de los datos. El número de validaciones cruzadas que utilizó el modelo fue 10, dado que fue donde se obtuvieron los mejores resultados. El modelo ocupa estas dos herramientas y el número de pre podas se fija en 40.

La pre poda, que ayuda a detener el árbol antes de que alcance el punto en que clasifica perfectamente los datos de entrenamiento, se fijo en 40. En palabras más exactas, es el número de nodos alternativos con los que cuenta la pre poda para evitar una división. La post poda, que permite que el modelo sobre ajuste a los datos y luego poda remplazando sub árboles por hojas, se activo. En este sentido no hay un parámetro que definir en el modelo, solo se puede activar o desactivar. Con estos últimos dos parámetros, los árboles resultaron mucho más compactos y con una predicción mucho mayor a árboles que contenían números de poda inferiores.

El tamaño mínimo para dividir en cada uno de los test que se realizan en los nodos, según el atributo seleccionado que entrega mayor información, se fijó en 15.

El tamaño mínimo de todas las hojas se fijó en 80. Es decir, el árbol clasifica hasta que todas las hojas terminales posean un número de registros mayor o igual a 80. El concepto por detrás de esta definición, se basa en que se desea obtener un árbol general y orientado grupos de clientes no tan pequeños. Es así, como los patrones y las reglas del árbol pueden estar enfocados hacia segmentos de clientes más amplios, donde las ofertas comerciales pueden ser más efectivas y concretas. Pues, no es la idea definir un conjunto de reglas para 5 clientes, ya que en términos prácticos no resulta útil.

La ganancia mínima de información, para cada división, se fijó en 0,1. Es decir, cada división por lo menos debe estar ganando un 10% de explicación a la incertidumbre generada hasta el momento. De esta manera la entropía del problema se va disminuyendo de manera veloz al tener una ganancia mínima de este orden. Nuevamente, lo que se busca es obtener un árbol compacto y general, de tal manera optimizar las ofertas comerciales.

La profundidad máxima del árbol se fijó en 50. La explicación de setear este parámetro en un número no tan pequeño es comprobar si es que los parámetros anteriormente explicados están cumpliendo su función.

El nivel de confianza se fijó en 0,25. Este es el valor por defecto y según la literatura genera buenos resultados en estos tipos de modelos. A grande rasgos, lo que dice este parámetro es que la clasificación errónea no puede ser mayor al 25% en cada nodo.

En la figura 39 se muestra la configuración del árbol de decisión.

Figura 39: Configuración de parámetros para el árbol de decisión.

Fuente: Elaboración Propia.

Al aplicar el modelo, con todos los parámetros mencionados con anterioridad y la ejecución de la validación cruzada, se obtuvieron los siguientes resultados, detallados en la figura 40.

Figura 40: Predicción del árbol SUS.

Correctos	13.248	83,39%
Incorrectos	2.639	16,61%
Total	15.887	100,00%

Fuente: Elaboración Propia.

Se puede notar que la predicción correcta corresponde a un 83,39% y que el error absoluto fue de un 16,61%. El resultado fue bastante exacto, pues con un 83,39% de probabilidad el modelo puede predecir la categoría de los clientes.

Los resultados entregados por el modelo fueron analizados por medio de una matriz de confusión, donde se compara la predicción con la categoría real de los clientes en estudio. En la figura 41 se detallan los resultados obtenidos. Las filas representan la predicción y las columnas la categoría real de los clientes.

Figura 41: Matriz de confusión árbol SUS.

		REAL		
		1	0	class precision
PREDICCIÓN	1	1.858	2.276	44,95%
	0	363	11.390	96,91%
	class recall	83,67%	83,34%	

Fuente: Elaboración Propia.

La predicción de los up-sellers se traduce en que con un 83,67% de probabilidad el modelo predice de manera correcta a los clientes que pueden realizar up-selling a Planes Multimedia. Y en caso opuesto, el modelo puede predecir con un 83,34% de probabilidad a los clientes que no realizarían up-selling.

Es necesario analizar los errores generados por el modelo. Para esto se deben considerar dos tipos de errores.

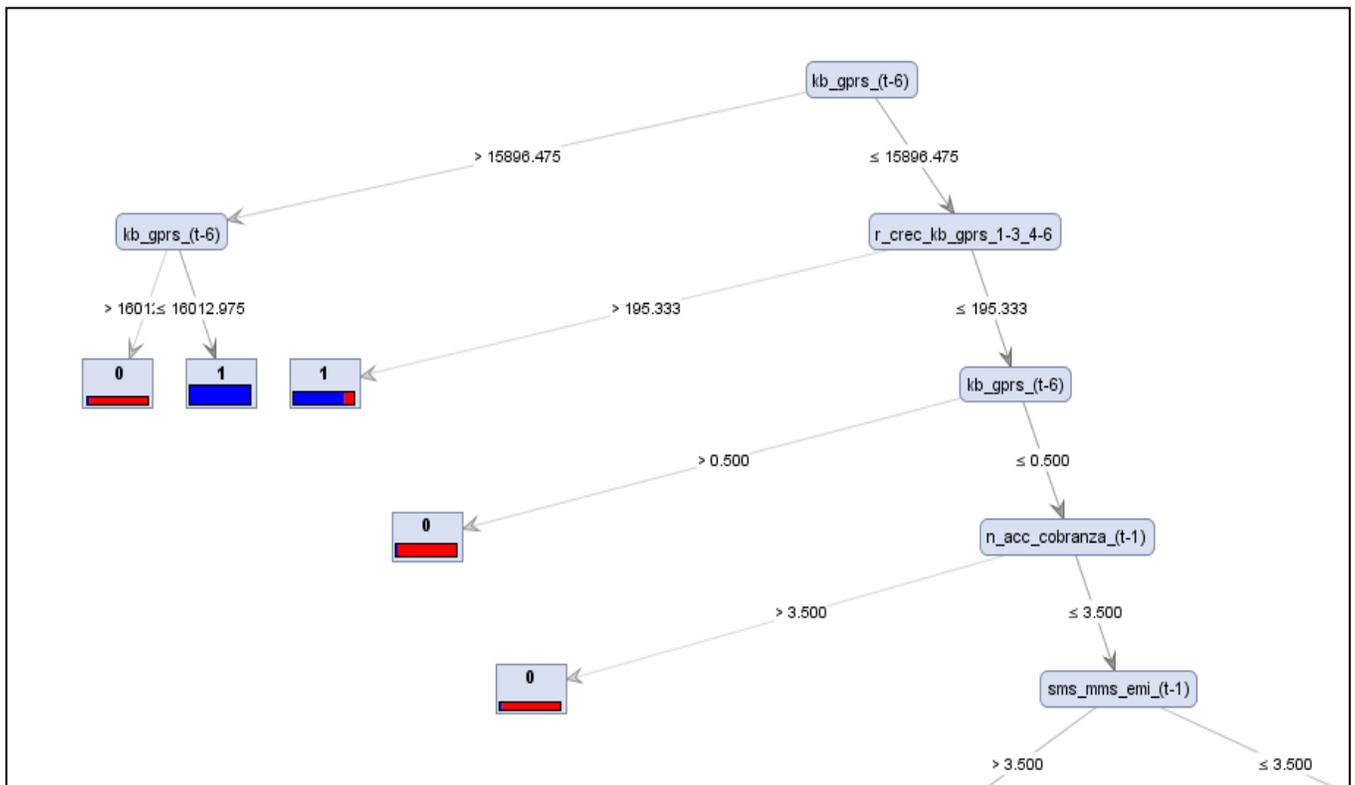
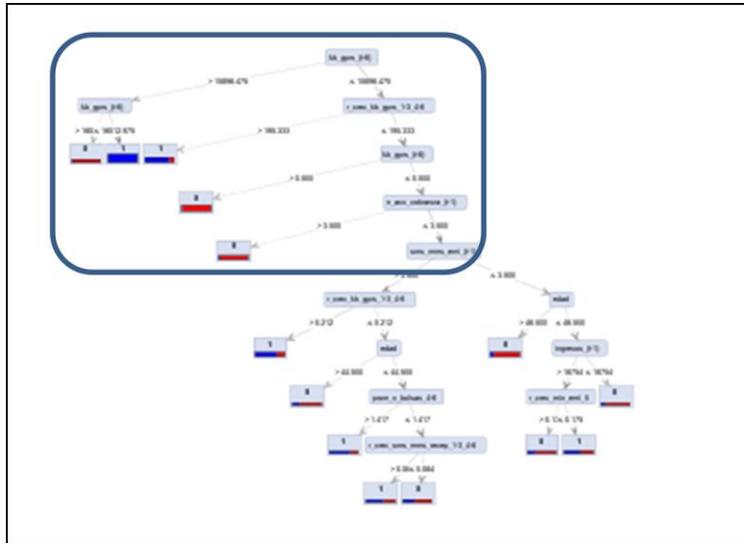
- Error de tipo 1: Corresponde al error que comete el modelo al decir que un cliente probablemente realice up-selling a un plan Multimedia, siendo que en la realidad el cliente no realizaría. Desde el punto de vista comercial, este error es más barato, ya que se les está ofreciendo el plan a clientes que no lo van a comprar. El resultado del ejercicio anterior, es que se otorgaría cierto esfuerzo a un cliente que no es propenso a comprar la oferta. Por lo tanto, este error sólo encarece la oferta comercial, ya que se invierte en algún cliente que no está dispuesto a aceptar la oferta. Finalmente el error de tipo 1 corresponde a un 14,6% ($\text{Error de tipo 1} = 2.276/15.524$), es decir, a un 14,6% de la muestra se le estaría ofreciendo un plan Multimedia que no lo contrataría.

- Error de tipo 2: Corresponde al error que comete el modelo al decir que un cliente no va a realizar up-selling a un plan Multimedia, siendo que en la realidad el cliente si lo realizaría. Desde el punto de vista comercial, este error es más caro, ya que no se le está ofreciendo la oferta al cliente que verdaderamente es propenso a comprar un Plan Multimedia. De esta manera se dejaría de percibir los ingresos que se generan al momento de que este cliente salte a un plan de mayor valor. Es decir, se pierde un cliente potencial que podría empezar a generar ingresos mayores para la compañía desde un periodo anterior. Finalmente el error de tipo 2 corresponde a un 2,6% ($\text{Error de tipo 2} = 363/13.611$), es decir, a un 2,6% de la muestra no se le estaría ofreciendo un plan Multimedia que si lo contrataría.

Después de todo el análisis con respecto a la calidad de la predicción, resulta necesario mostrar el modelo, de tal manera de poder observar cuales son los atributos

más explicativos y los patrones que caracterizan a los clientes que realizan up-selling. En la figura 42 se presenta el árbol de decisión generado.

Figura 42: Modelo de árbol de decisión para clientes SUS.



número de acciones de cobranza, mensajes emitidos, la edad, la facturación del cliente (equivalente a ingresos para la compañía), el promedio de número de bolsas de servicios contratadas y la tasa de crecimiento de mensajes recibidos.

Según la información entregada por el modelo, es posible caracterizar a aquellos clientes que son propensos a realizar el up-selling. De esta manera se pueden observar reglas generales. Para ello se define como regla general a hojas terminales que poseen más de 60 clientes de la muestra y la proporción que contrataría un plan Multimedia fuese mayor al 55%.

- Los clientes cuya descarga de datos de Internet se encuentra entre los 15.896 kb y los 16.012 kb, la proporción de cliente que contrataría un plan Multimedia será el 100%, que equivale a 664 clientes.

- Los clientes cuya descarga de datos sea menor a 15.896 kb y su tasa de crecimiento trimestral de descarga de datos sea mayor al 195%, la proporción de clientes que contrataría un plan Multimedia corresponde al 83,15%, que equivale a 306 clientes.

- Los clientes cuya tasa de crecimiento trimestral sea mayor al 195%, su descarga de datos sea menor a 0.5 kb, el número de cobranza que se le aplica sea menor a 3.5, sus mensajes emitidos sean menores a 3.5, sean menores de 46 años, su facturación sea mayor a \$16.754 y su tasa de crecimiento de minutos sea menor o igual a 17,5%, la proporción de clientes que contrataría un plan Multimedia corresponde al 56,29%, que equivale a 76 clientes.

- Los clientes cuya tasa de crecimiento trimestral sea mayor al 195%, su descarga de datos sea menor a 0.5 kb, el número de cobranza que se le aplica sea menor a 3.5, sus mensajes emitidos sean mayores a 3.5 y su tasa de crecimiento trimestral de descarga de datos sea mayor al 21,2%, la proporción de clientes que contrataría un plan Multimedia corresponde al 73,62%, que equivale a 201 clientes.

- Los clientes cuya tasa de crecimiento trimestral sea mayor al 195%, su descarga de datos sea menor a 0.5 kb, el número de cobranza que se le aplica sea menor a 3.5, sus mensajes emitidos sean mayores a 3.5, su tasa de crecimiento trimestral de descarga de datos sea menor al 21,2%, su edad sea menor a 44 años y su promedio de número de bolsas de servicios contratadas sea mayor a 1,4 la proporción de clientes que contrataría un plan Multimedia corresponde al 71,91%, que equivale a 64 clientes.

- Los clientes cuya tasa de crecimiento trimestral sea mayor al 195%, su descarga de datos sea menor a 0.5 kb, el número de cobranza que se le aplica sea menor a 3.5,

sus mensajes emitidos sean mayores a 3.5, su tasa de crecimiento trimestral de descarga de datos sea menor al 21,2%, su edad sea menor a 44 años, su promedio de número de bolsas de servicios contratadas sea menor o igual a 1,4 y su tasa de crecimiento trimestral de mensajes recibidos sea mayor a 6,4%, la proporción de clientes que contrataría un plan Multimedia corresponde al 59,25%, que equivale a 64 clientes.

Con esta información es posible generar ofertas focalizadas a segmentos de clientes donde la propensión de compra a planes Multimedia es elevada. Se pueden generar ofertas comerciales a aquellos clientes con consumo de descarga de datos de internet, que no tengas problemas financieros con la compañía, que trafique mensajes de texto o mensajes multimedia, que además sea menor de 46 años, que su facturación mensual sea superior a los \$16.000, que su tasa de crecimiento de minutos sea elevada, que esté recibiendo más mensajes multimedia en el último tiempo y que esté contratando bolsas de distintos servicios para suplir la falta de servicios en su plan.

8.5.2 APLICACIÓN DE ÁRBOL DE DECISIÓN A CLIENTES CC

El desarrollo de la aplicación del modelo de árbol para clientes Cuenta Controlada es análoga a la descrita para los clientes SUS. Se siguió la misma estructura y a continuación se muestra su desarrollo.

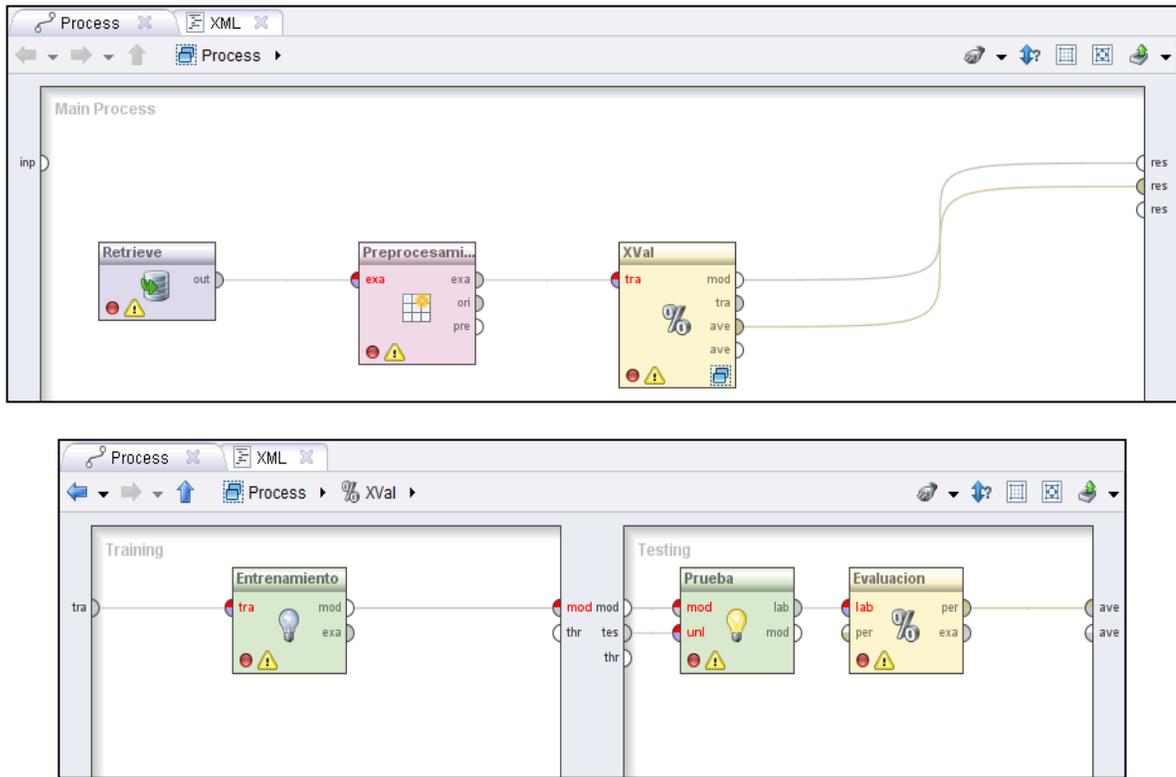
En el conjunto de datos objetivo de los clientes Cuenta Controlada, la cantidad de clientes up-seller es de 4,5% del total. Por ende, el restante 95,5% corresponden a clientes que se mantuvieron en planes distintos a los planes Multimedia. Con estos datos es posible apreciar que el conjunto final de datos está totalmente desbalanceado.

La solución de desarrollo para este problema es idéntica a la de SUS, por lo tanto se dividió la muestra total en Muestra de Entrenamiento y Muestra de Prueba. La muestra de entrenamiento contiene el 70% de los datos de la muestra total y su contenido es proporcional entre clientes que realizaron up-selling y clientes que aún no lo han realizado (50-50). Para el caso de la muestra de prueba, posee el 30% de los datos y su proporción de clientes up-sellers y no up-seller vuelve a ser la real, es decir, 4,5% para los up-sellers y 95,5% para el resto.

El árbol de decisión se creó utilizando todos los atributos que caracterizan cada uno de los clientes en estudio. La variable independiente de este modelo vuelve a ser “**es_pt(t)**” que corresponde a “1” si es up-seller y “0” si no lo es.

El proceso comienza con ingresar los datos objetivos como input, luego realizar el procesamiento de limpieza de datos para rellenar los valores perdidos y por último la validación cruzada que aplica el árbol de decisión.

Figura 43: Proceso de aplicación de árbol para clientes CC.



Fuente: Elaboración Propia.

Se corroboró a través del método de “Bootstrapping” la distribución del error en el muestreo de entrenamiento. Esto significa que se crearon 10 muestras de entrenamiento aleatorias para aproximar la varianza de la predicción de los modelos. Entregando una varianza de 1,63%, lo que indica que el modelo es confiable para distintas muestras de datos.

Del mismo modo que para los clientes SUS, se aplicaron distintos tipos de árboles con distintos tipos de parámetros hasta llegar al árbol que entregara los mejores resultados de predicción. El criterio de selección de atributos y la configuración de parámetros para el árbol es idéntica que para el caso de los clientes SUS.

En la figura 44 se muestra la configuración del árbol de decisión.

Figura 44: Configuración del árbol para clientes CC.

The image shows a configuration window titled "Entrenamiento (Decision Tree)". It contains the following settings:

- criterion: information_gain
- minimal size for split: 15
- minimal leaf size: 80
- minimal gain: 0.1
- maximal depth: 50
- confidence: 0.25
- number of prepruning altern...: 40
- no pre pruning
- no pruning

Fuente: Elaboración Propia.

Al aplicar el modelo se obtuvieron los siguientes resultados, detallados en la figura 45.

Figura 45: Predicción árbol CC.

Correctos	17.498	75,50%
Incorrectos	5.679	24,50%
Total	23.178	100,00%

Fuente: Elaboración Propia.

Se puede notar que la predicción correcta equivale a un 75,5% y que el error absoluto corresponde a un 24,5%. El resultado obtenido por el modelo se considera relativamente bueno, ya que se espera que por lo menos la predicción correcta este sobre el 80%. Pero sin duda que agrega bastante información para el análisis y la generación de ofertas comerciales.

En la figura 46 se muestra la matriz de confusión entregada por el modelo.

Figura 46: Matriz de confusión árbol CC.

		REAL		
		1	0	class precision
PREDICCIÓN	1	822	5.458	13,08%
	0	221	16.677	98,69%
class recall		78,77%	75,34%	

Fuente: Elaboración Propia.

La predicción de los up-sellers se estima en que con un 78,77% de probabilidad el modelo es capaz de identificar que el cliente realiza el up-selling. Para el caso de los de los clientes que no realizan up-selling el modelo estima con un 75,34% de probabilidad su comportamiento.

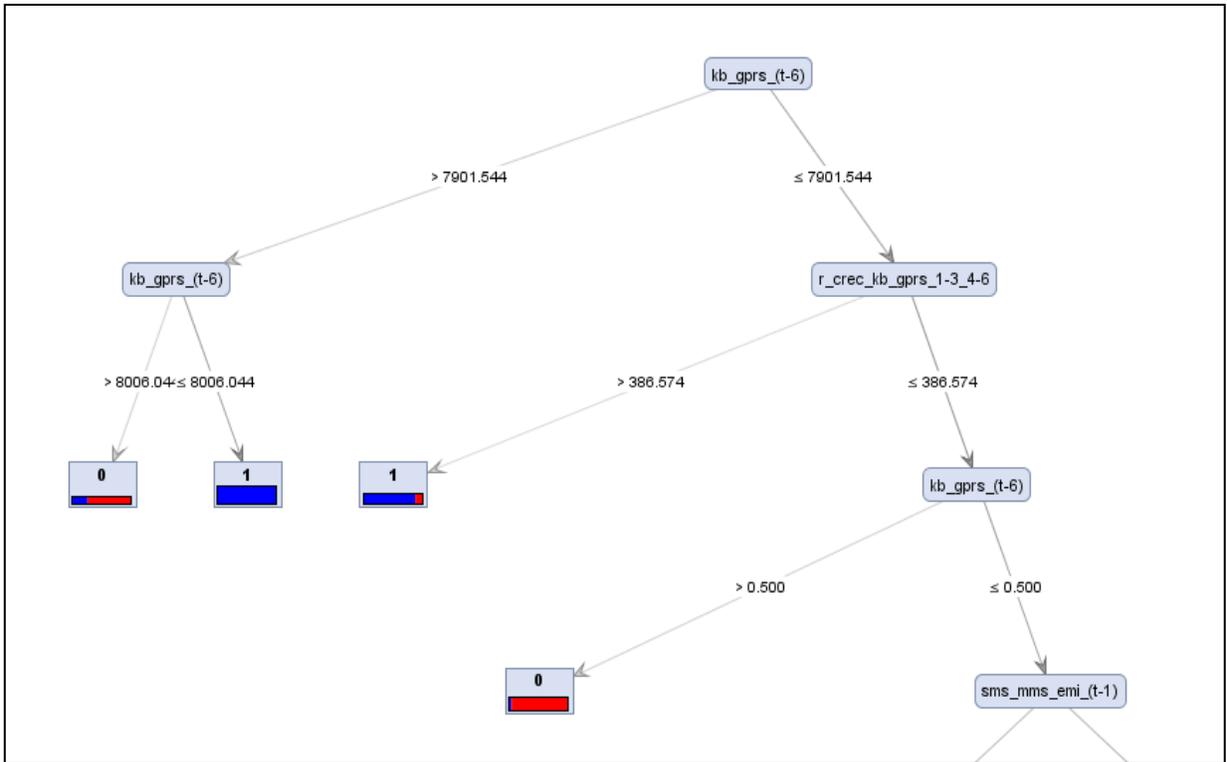
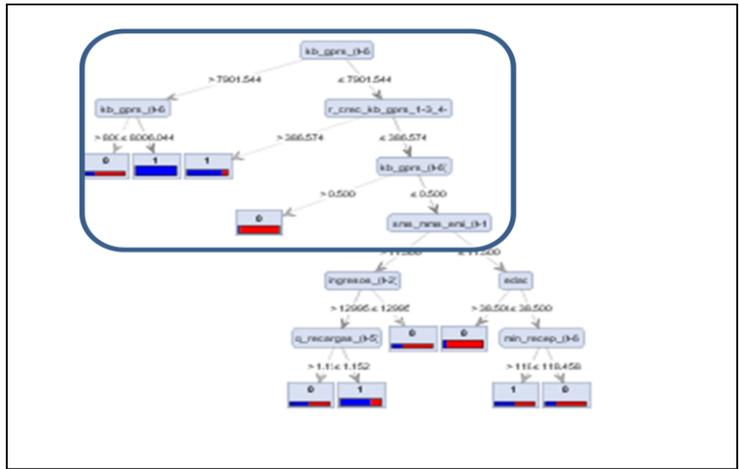
Para analizar de manera correcta los errores, se indagó en los errores de tipo 1 y de tipo 2.

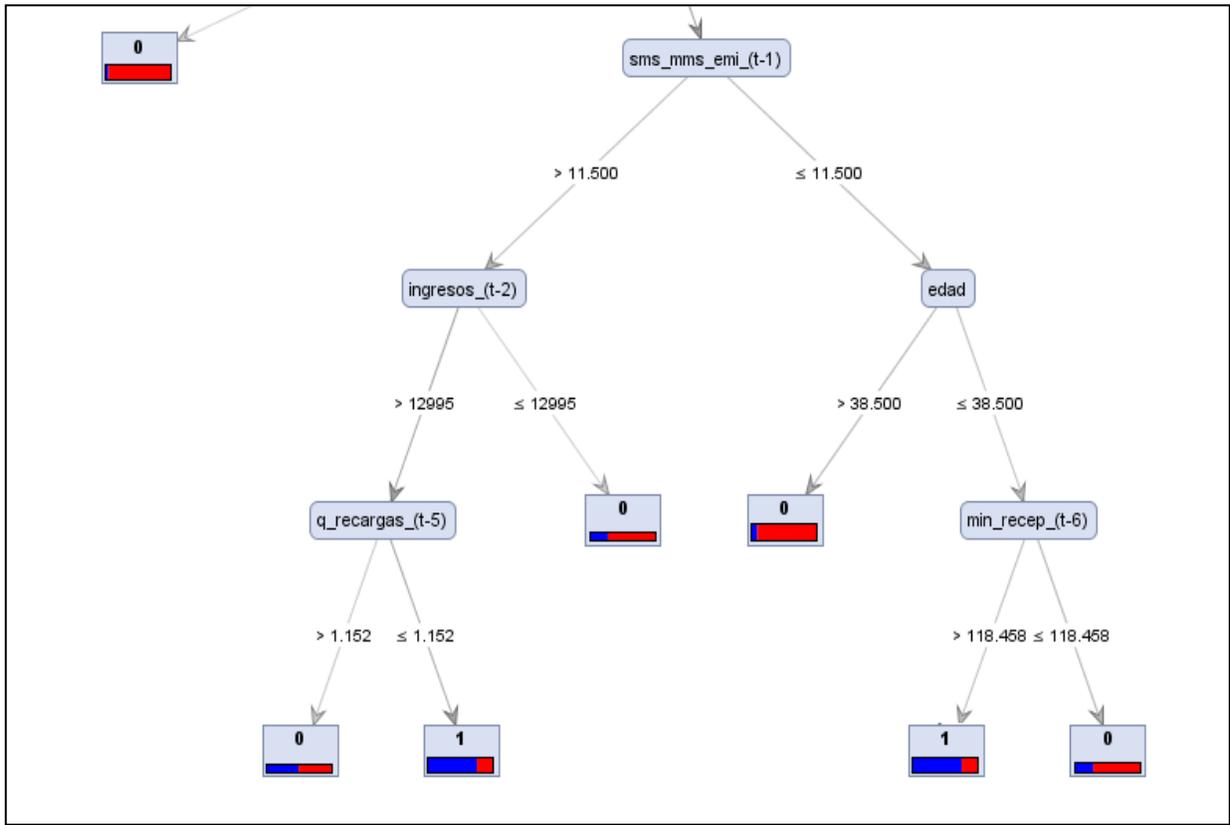
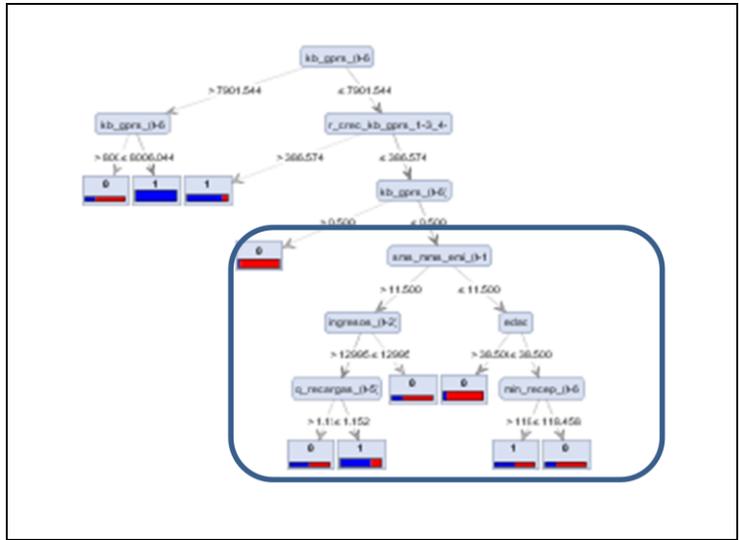
- Error de tipo 1: El error de tipo 1 corresponde a un 23,7% (Error de tipo 1 = $5.458/22.957$), es decir, a un 23,7% de la muestra se le estaría ofreciendo un plan Multimedia que no lo contrataría. Considerando que el costo de este error no es económicamente tan grave, el resultado regular, ya que se estaría gastando esfuerzo por clientes que no son propensos al up-selling. Pues, el modelo solo se equivocaría con un 23,7% de probabilidad

- Error de tipo 2: El error de tipo 2 corresponde a un 1,24% (Error de tipo 2 = $221/17.720$), es decir, a un 1,24% de la muestra no se le estaría ofreciendo un plan Multimedia que si lo contrataría. Al ser este error más caro económicamente que el error de tipo 1, se espera sea menor y exitosamente es así. Solo se comete este error con un 1,24% de probabilidad.

Luego del análisis correspondiente a la predicción, resulta útil mostrar el modelo generado, de tal manera de poder identificar cuáles son los atributos que explican el salto de un plan no target a un plan Multimedia. En la figura 47 se presenta el árbol de decisión.

Figura 47: Modelo del árbol para clientes CC.





Fuente: Elaboración Propia.

Al observar el modelo se puede apreciar que la variable más efectiva usando el criterio de Ganancia de Información son los kilobytes descargados. En segundo nivel vuelve a aparecer la descarga de datos y la tasa de crecimiento trimestral de descarga de datos. En nodos más ramificados se puede encontrar los mensajes de texto y

mensajes multimedia emitidos, la facturación, la edad, la cantidad de recargas y los minutos recibidos.

Según la información entregada por el modelo, es posible caracterizar a aquellos clientes que son propensos a realizar el up-selling. Para ello se define como regla general a hojas terminales que poseen más de 60 clientes de la muestra y la proporción que contrataría un plan Multimedia fuese mayor al 55%.

- Los clientes cuya descarga de datos de internet se encuentre entre los 7.901 kb y los 8.006 kb, la proporción de clientes que contrataría un plan Multimedia será del 100%, que equivale a 291 clientes.

- Los clientes cuya descarga de datos sea inferior a 7.901 kb y su tasa de crecimiento trimestral de descarga de datos sea mayor a 386,5%, la proporción de clientes que contrataría un plan Multimedia sería de 86,98%, que equivale a 127 clientes.

- Los clientes cuya tasa de crecimiento trimestral de descarga de datos sea menor o igual a 386,5%, su descarga de datos sea inferior a 0.5 kb, sus mensajes de texto y multimedia, en su conjunto, sean mayores a 11.5, su facturación sea mayor a \$12.995 y su cantidad de recargas sea menor o igual a 1,15, la proporción de clientes que contrataría un plan Multimedia sería de 74,86%, que equivale a 134 clientes.

- Los clientes cuya tasa de crecimiento trimestral de descarga de datos sea menor o igual a 386,5%, su descarga de datos sea inferior a 0.5 kb, sus mensajes de texto y multimedia, en su conjunto, sean menores o iguales a 11.5, su edad sea menor a 38 años, sus minutos recibidos sean mayores a 118,45, la proporción de clientes que contrataría un plan Multimedia sería de 66,94%, que equivale a 81 clientes.

Con la información arrojada por el modelo es posible generar ofertas focalizadas a segmentos de clientes donde la propensión de compra a planes Multimedia es, casi, segura. En el caso de los clientes CC, que son propensos a realizar up-selling se caracterizan por descargar de datos de internet, son clientes que trafiquen mensajes de texto multimedia, que su facturación es superior a los \$13.000, la cantidad de recargas de servicios de voz es menor a 1,52, que sean clientes menores de 38 años y que este recibiendo un alta proporción de minutos recibidos.

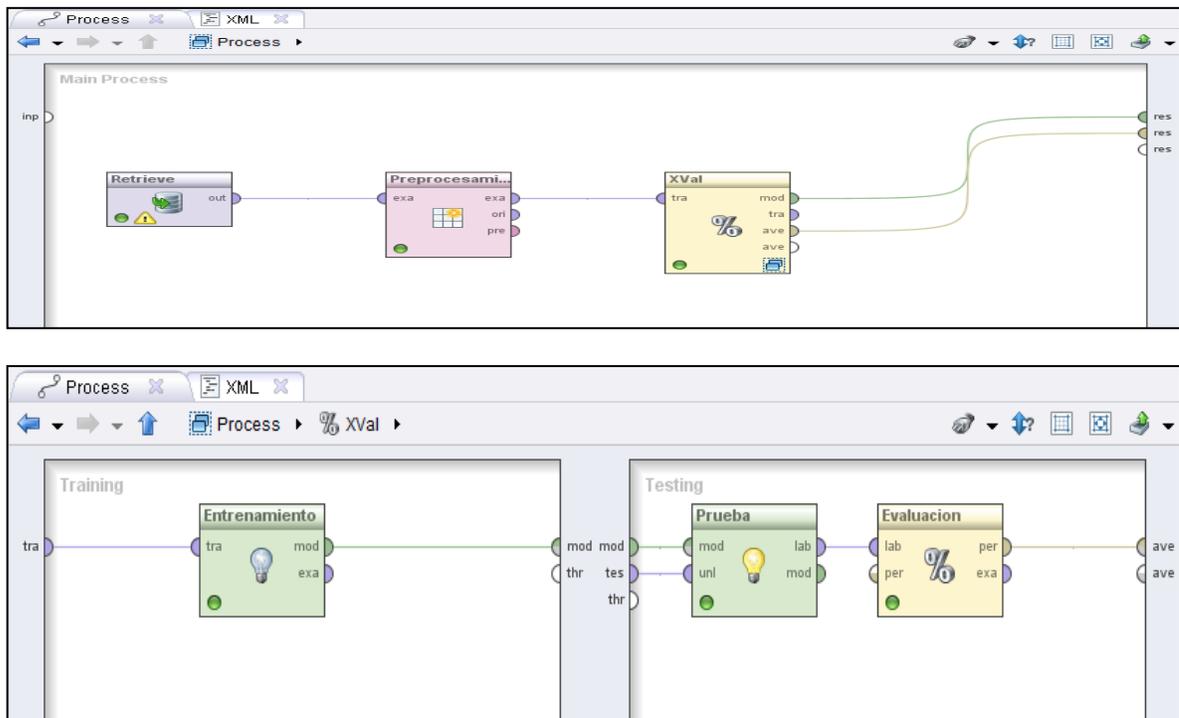
8.5.3 APLICACIÓN DE RED NEURONAL A CLIENTES SUS

Se aplicó otra técnica de minería de datos con el objetivo de predecir a través de otra alternativa los clientes que son propensos a comprar un plan Multimedia. De esta manera es posible competir en predicción con el resto de las técnicas aplicadas.

Este modelo se realizó con la misma metodología que el árbol de decisión, es decir, se ocuparon las dos muestras (entrenamiento y prueba) ya que los datos estaban desbalanceados según la categorización de los clientes. La única salvedad es que no se tomaron en cuenta las variables categóricas, solo se utilizaron las variables numéricas. El motivo de esta determinación es que la red ocupada bajo la herramienta RapidMiner sólo permite ocupar atributos numéricos. Por lo tanto, quedaron fuera de conjunto de atributos a ocupar el género, sector económico, actividad, equipo terminal, tecnología del equipo terminal, entre otros.

El proceso de parte con la entrada de los datos objetivo, luego con el procesamiento de limpieza orientado a los valores perdidos y por último la validación cruzada donde se aplica la red neuronal. En la figura 48 se muestra el proceso.

Figura 48: Proceso de red neuronal para clientes SUS.



Fuente: Elaboración Propia.

La aplicación de la red neuronal estuvo caracterizada por realizar varias pruebas con distintas configuraciones en los parámetros de la red. Se escogió la configuración que reportó mejores resultados.

En primer lugar, se decidió crear tres capas ocultas para el modelo. Se fijó esta cantidad de capas ocultas ya que entregó los mejores resultados predictivos.

El número de ciclos utilizado para la formación de la red se fijó en 500, de esta manera se evita el sobre entrenamiento de la red.

La tasa de aprendizaje, que define cuanto cambian los pesos definidos en cada etapa, se fijó en 0,3. Es decir, no puede cambiar su valor más allá del 30% por pasos siguientes en la red.

El momentum se fijó en 0,2. Lo que produce es aumentar una fracción de la actualización de los pesos, de la etapa anterior a la actual, pues de esta manera se puede evitar máximos locales y las direcciones de optimización se suavizan.

Se activó el parámetro que permite indicar si es que los datos de entrada deben ser mezclados antes de comenzar el aprendizaje del modelo.

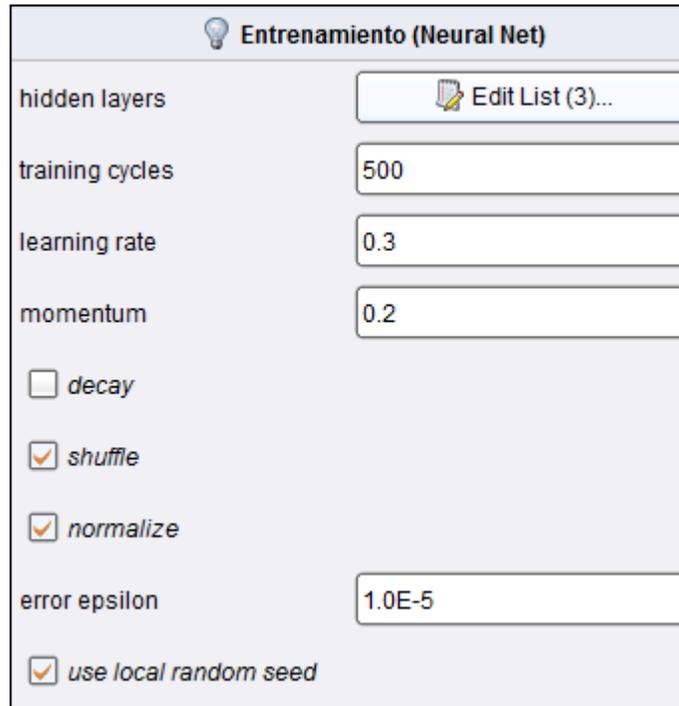
Se activó la normalización, que indica si los datos de entrada deben ser normalizados entre -1 y +1.

El épsilon de error se fijó en 0,00005, donde la optimización se detiene si el error de entrenamiento baja de este valor.

Se activó el uso de una semilla aleatoria local. De esta manera se aseguro que la secuencia de valores aleatorios para inicializar los pesos de la red no fuera diferente cada vez que el método fuese ejecutado. Pues así, no se crean distintos métodos para un mismo conjunto de datos.

En la figura 49 se muestra la configuración de la red neuronal.

Figura 49: Configuración de la red neuronal para clientes SUS.



Entrenamiento (Neural Net)

hidden layers Edit List (3)...

training cycles

learning rate

momentum

decay

shuffle

normalize

error epsilon

use local random seed

Fuente: Elaboración Propia.

Al aplicar el modelo con la configuración detallada con anterioridad, se obtuvieron los siguientes resultados, detallados en la tabla 51.

Figura 50: Predicción red neuronal SUS.

Correctos	10.149	63,88%
Incorrectos	5.738	36,12%
Total	15.887	100,00%

Fuente: Elaboración Propia.

Se puede apreciar que la predicción correcta corresponde a un 63,88% y que el error absoluto fue de 36,12%. El resultado no es del todo esperanzador, ya que la predicción es muy baja.

Los matriz de confusión, entregada por el modelo se puede apreciar en la figura 51.

Figura 51: Matriz de confusión red neuronal SUS.

		REAL		
		1	0	class precision
PREDICCIÓN	1	1.668	5.185	24,34%
	0	553	8.481	93,88%
	class recall	75,11%	62,06%	

Fuente: Elaboración Propia.

La predicción de los up-sellers se traduce en que con un 75,11% de probabilidad el modelo predice de manera correcta a los clientes que pueden realizar up-selling a Planes Multimedia. Y en caso opuesto, el modelo puede predecir con un 62,06% de probabilidad a los clientes que no realizarían up-selling. Es posible notar que la predicción de los que realmente realizaron up-selling no es mala pero la predicción de los clientes que no realizan up-selling no es buena.

Es necesario analizar los errores generados por el modelo. Para esto se deben considerar dos tipos de errores.

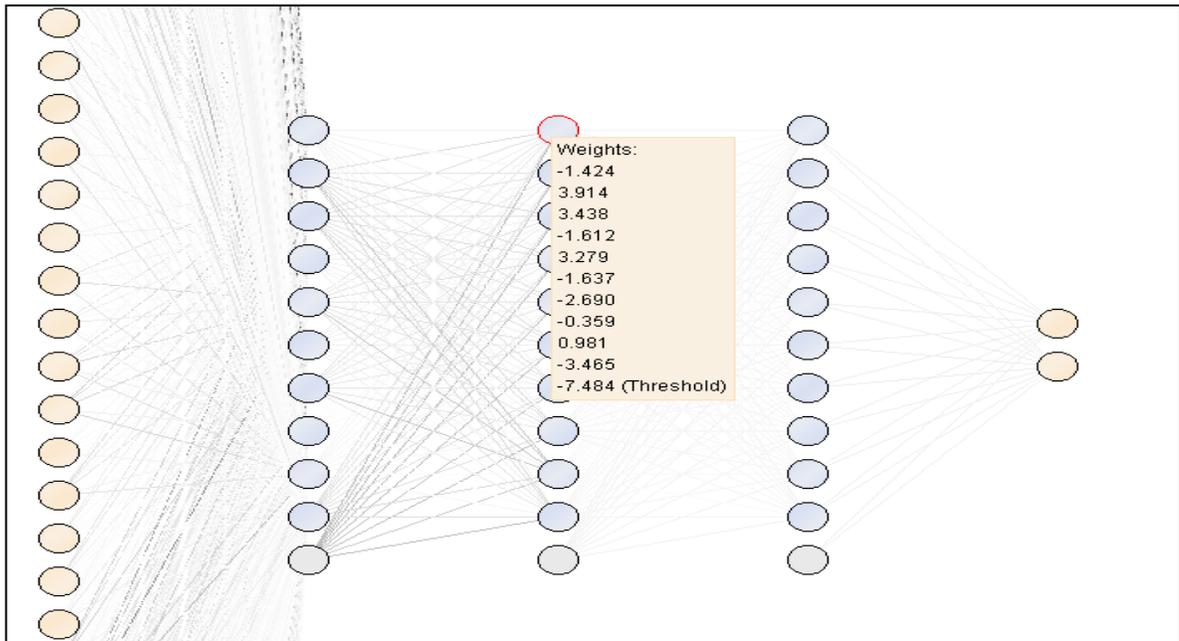
- Error de tipo 1: El error de tipo 1 corresponde a un 33,8% (Error de tipo 1 = $5.185/15.334$). Si bien el error se puede considerar relativamente alto, el error de tipo 1 es considerado menos costo comercialmente, por lo que el impacto no es del todo grave. En este sentido, con un 33,8% de probabilidad se le ofrecería un Plan Multimedia a un cliente que no es propenso a comprarlo.

- Error de tipo 2: El error de tipo 2 corresponde a un 5,16% (Error de tipo 2 = $553/10.702$). Considerando que este error es más caro comercialmente, se necesita que no sea tan elevado. Afortunadamente, el error de tipo 2 es menor que el error de tipo 1 y con un 5,16% de probabilidad no se le ofrecerá el Plan Multimedia a un cliente que si lo compraría.

En conclusión, la predicción del modelo no es del todo correcta pero el error de tipo 2, que se considera el más caro, es bastante bajo. Por lo tanto, mirando el modelo desde una perspectiva comercial el modelo sirve para generar ofertas focalizadas a segmentos de clientes.

En la figura 52 se muestra parte de la red generada.

Figura 52: Modelo red neuronal SUS.



Fuente: Elaboración Propia.

8.5.4 APLICACIÓN DE RED NEURONAL A CLIENTES CC

Para el caso de los clientes CC, se aplico la red neuronales con el mismo proceso y con la misma configuración de los parámetros de la red que para los clientes SUS.

Al aplicar el modelo, se obtuvieron los resultados que detalla la figura 53.

Figura 53: Predicción red neuronal CC.

Correctos	15.477	66,78%
Incorrectos	7.701	33,22%
Total	23178	100,00%

Fuente: Elaboración Propia.

Se puede apreciar que la predicción correcta equivale a un 66,78% y que el error absoluto corresponde a un 33,22%. Si bien el resultado no es muy bueno, es de vital importancia analizar los costes de los errores.

La matriz de confusión entregada por el modelo, se puede observar en la figura 54.

Figura 54: Predicción red neuronal CC.

		REAL		
		1	0	class precision
PREDICCIÓN	1	802	7.459	9,70%
	0	241	14.676	98,38%
class recall		76,85%	66,30%	

Fuente: Elaboración Propia.

La predicción de los up-sellers se traduce en que con un 76,85% de probabilidad el modelo predice de manera correcta a los clientes que pueden realizar up-selling a Planes Multimedia. Y en caso opuesto, el modelo puede predecir con un 66,30% de probabilidad a los clientes que no realizarían up-selling. Es posible notar que la predicción de los que realmente realizaron up-selling no es mala, pero la predicción de los clientes que no realizan up-selling no del todo exacta.

Es necesario analizar los errores generados por el modelo. Para esto se deben considerar dos tipos de errores.

- Error de tipo 1: El error de tipo 1 corresponde a un 32,5% (Error de tipo 1 = $7.459/22.937$), es decir, a un 32,5% de la muestra se le estaría ofreciendo un plan Multimedia que no lo contrataría.

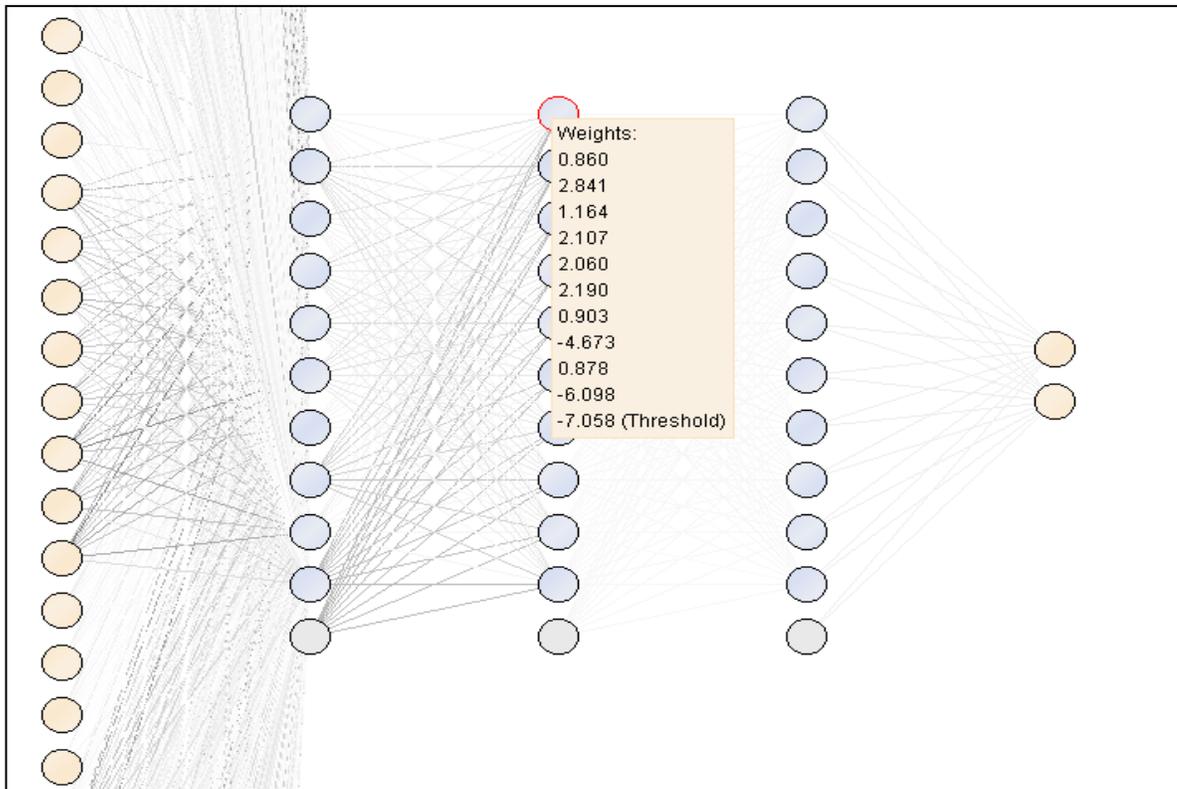
- Error de tipo 2: El error de tipo 2 corresponde a un 1,53% (Error de tipo 2 = $241/15.719$), es decir, a un 1,53% de la muestra no se le estaría ofreciendo un plan Multimedia que si lo contrataría. El error es muy bajo, por lo tanto no se estarían perdiendo ventas con este modelo.

En conclusión, la predicción no es del todo correcta, pero el error de tipo 2, que se considera el más caro, es muy bueno. Por lo tanto, mirando el modelo desde una perspectiva comercial el modelo sirve para generar ofertas focalizadas a segmentos de clientes.

A diferencia de los árboles de decisión, las redes neuronales no aportan mucho conocimiento e información para interpretar los resultados, ya que se entrega una estructura compleja. De esta manera, la red neuronal se considera como una caja negra donde se entrega un input (los datos) y se ve sólo el output (la predicción) sin poder analizar el comportamiento de la predicción. La agregación de valor y conocimiento para caracterizar a los clientes que podrían llegar a realizar up-selling es nula.

En la figura 55 se muestra parte de la red generada.

Figura 55: Modelo red neuronal CC.



Fuente: Elaboración Propia.

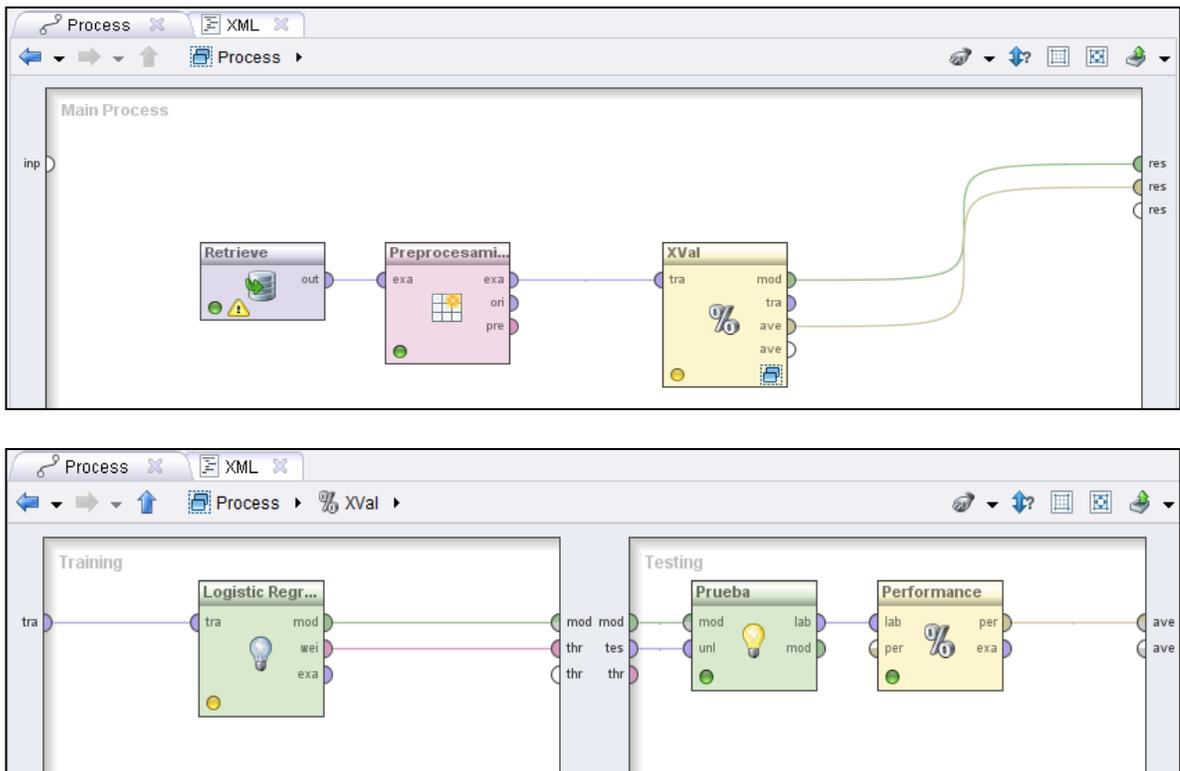
8.5.5 APLICACIÓN DE REGRESIÓN LOGÍSTICA A CLIENTES SUS

Por último, se aplicó una tercera herramienta de minería de datos con el objetivo de comparar las tres técnicas y elegir la que entregue los mejores resultados.

Para generar la regresión logística se aplicó la misma metodología que en las técnicas anteriores, es decir, se ocupó la muestra de entrenamiento y la muestra de prueba. Y al igual que en la red neuronal, solo se escogieron los atributos numéricos, ya que la herramienta de minería de datos no permite trabajar las regresiones con atributos categóricos.

El proceso es similar a los anteriores, comienza con la entrada de los datos objetivos, luego se realiza el procesamiento de limpieza de valores perdidos y por último se realiza la validación cruzada donde se aplica la regresión logística. En la figura 56 se detalla el proceso.

Figura 56: Proceso regresión logística para clientes SUS.



Fuente: Elaboración Propia.

La aplicación de la regresión logística se caracterizó por probar con distintas configuraciones de parámetros hasta llegar a la configuración final que otorgara los mejores resultados en términos de predicción.

El tipo de kernel elegido fue DOT, ya que entregó los mejores resultados. Cabe destacar que el tipo de kernel permite utilizar versiones lineales de algoritmos no lineales.

El primer parámetro que se definió fue la complejidad de la regresión. Para este caso se fijó en 1, ya que entregó los mejores resultados.

El ϵ de convergencia, que entrega la precisión de las condiciones de optimización de KKT, se fijó en 0,001. El motivo de la elección, al igual que en parámetros anteriores se debe a que en la acción de prueba y error, fue el valor que entregó los mejores resultados.

El máximo de iteraciones se fijó en 100000, valor bastante alto para que la regresión pueda iterar la cantidad de veces necesarias y encontrar los mejores pesos para los atributos. Al colocar valores mayores a 100000, el tiempo de ejecución aumentaba exponencialmente.

En la figura 57 se muestra la configuración de la regresión logística.

Figura 57: Configuración de parámetros regresión logística para clientes SUS.

Fuente: Elaboración Propia.

Al aplicar el modelo, se obtuvieron los resultados mostrados en la tabla 59.

Figura 58: Predicción regresión logística para clientes SUS.

Correctos	10.664	67,12%
Incorrectos	5.223	32,88%
Total	15.887	100,00%

Fuente: Elaboración Propia.

La predicción del modelo es 67,12% y el error absoluto corresponde a un 32,88%. Es posible apreciar que la predicción no es mejor que los modelos anteriores y si bien entrega cierto grado de información no es del todo confiable, ya que la cantidad de predicción incorrecta es muy elevada.

La matriz de confusión generada por el modelo se puede observar en la figura 59.

Figura 59: Matriz de confusión regresión logística para clientes SUS.

		REAL		
		1	0	class precision
PREDICCIÓN	1	1.471	4.473	24,75%
	0	750	9.193	92,46%
class recall		66,24%	67,27%	

Fuente: Elaboración Propia.

La predicción de los up-sellers es de un 66,24% y la predicción de los no up-sellers es de un 67,27%. En este sentido, los dos enfoques no tienen una predicción confiable.

A continuación se detalla el análisis de los tipos de errores.

- Error de tipo 1: El error de tipo 1 corresponde a un 29,5% (Error de tipo 1 = 4.473/15.137), es decir, a un 29,5% de la muestra se le estaría ofreciendo un plan Multimedia que no lo contrataría.

- Error de tipo 2: El error de tipo 2 corresponde a un 6,6% (Error de tipo 2 = 750/11.414), es decir, a un 6,6% de la muestra no se le estaría ofreciendo un plan Multimedia que si lo contrataría. El error es muy bajo y al ser este error más costoso económicamente, el modelo genera muy buenos resultados.

La predicción de la regresión logística no entregó los resultados apropiados y estuvo por debajo de los modelos generados con anterioridad. En el Anexo 14 es posible observar los pesos de cada uno de los atributos del modelo.

8.5.6 APLICACIÓN DE REGRESIÓN LOGÍSTICA A CLIENTES CC

Para el caso de la regresión logística de cliente CC, se aplicó de forma análoga a los clientes SUS la metodología, el proceso y la configuración de los parámetros.

Los resultados del modelo se pueden observar en la figura 60.

Figura 60: Predicción regresión logística para clientes CC.

Correctos	12.600	54,36%
Incorrectos	10.578	45,64%
Total	23.178	100,00%

Fuente: Elaboración Propia.

Asimismo, se obtuvo la matriz de confusión que se muestra en la figura 61.

Figura 61: Matriz de confusión regresión logística para clientes SUS.

		REAL		class precision
		1	0	
PREDICCIÓN	1	804	10.340	7,22%
	0	239	11.795	98,02%
class recall		77,12%	53,29%	

Fuente: Elaboración Propia.

La predicción de los up-sellers corresponde a un 77,12% y la predicción de los no up-sellers es de un 53,29%. En este sentido, sólo la predicción de los up-sellers es confiable, ya que la predicción de los clientes que no realizan up-selling es cercana al 50% de probabilidad, misma probabilidad que se tiene al lanzar una moneda y otorgar la categoría de esta forma.

A continuación se detalla el análisis de los tipos de errores.

- Error de tipo 1: El error de tipo 1 corresponde a un 45% (Error de tipo 1 = $10.340/22.939$), es decir, a un 45% de la muestra se le estaría ofreciendo un plan Multimedia que no lo contrataría. Si bien este error es menos costoso, la probabilidad de error es bastante alta, por lo tanto el costo aumenta en comparación al resto de los modelos.

- Error de tipo 2: Finalmente el error de tipo 2 corresponde a un 1,8% (Error de tipo 2 = $239/12.838$), es decir, a un 1,8% de la muestra no se le estaría ofreciendo un plan Multimedia que si lo contrataría.

La predicción de la regresión logística no entregó los resultados apropiados y estuvo por debajo de los modelos generados con anterioridad. La única salvedad es que el error más costoso fue el que contiene un menor error y la predicción para los clientes que realizan up-selling es bastante buena, por este motivo la generación de ofertas focalizadas a segmentos de clientes puede ser que no sea tan costosa en términos económicos. En el Anexo 15 es posible observar los pesos de cada uno de los atributos del modelo.

En conclusión y observando los resultados anteriormente expuestos, los modelos generados tienen la característica de poseer un error de tipo dos bastante bajo. Esto implica que existe muy baja probabilidad de que no se le ofrezca un Plan Multimedia a alguien que es propenso a comprar. Dado que el error de tipo dos es similar en todos los modelos, lo que se debe comparar es la predicción general del modelo, en este sentido el árbol de decisión es el que mejor predice el up-selling, tanto para clientes Suscripción como para clientes Cuenta Controlada. Además, es el modelo que genera mayor ganancia de información y conocimiento intuitivo para poder comprender ofertas focalizadas para segmentos de clientes.

8.6 EVALUACIÓN ECONÓMICA

Esta evaluación económica se utilizará los resultados obtenidos a través de los árboles de decisiones, tanto para SUS como para CC. A continuación se presentan las matrices de confusión en la figura 62 y la figura 63 respectivamente.

Figura 62: Matriz de confusión árbol de decisión para clientes SUS.

	1	0	total
1	1858	2276	4134
0	363	11390	11753
total	2221	13666	

Fuente: Elaboración Propia.

Figura 63: Matriz de confusión árbol de decisión para clientes CC.

	1	0	total
1	822	5458	6280
0	221	16677	16898
total	1043	22135	

Fuente: Elaboración Propia.

Lo que se pretende realizar es medir la ganancia en términos de valor futuro del cliente al adelantar el cambio de plan hacia un Plan Multimedia. Si un cliente realiza el up-selling estará destinando una mayor cantidad de dinero hacia la compañía de la que facturaba con el plan antiguo. Por lo tanto, la facturación del clientes se distingue entre aporte con Plan Multimedia o sin Plan Multimedia. Y además, se debe aplicar la diferencia entre clientes SUS y CC. De esta manera se define la rentabilidad promedio que generan los clientes a la compañía con los distintos planes.

Figura 64: Rentabilidad para clientes SUS y CC.

Posee Plan Multimedia	SUS	CC
NO	\$ 19.000	\$ 13.000
SI	\$ 31.773	\$ 18.990

Fuente: Elaboración Propia.

Según la recomendación de expertos de la compañía, el VPN total de un cliente se debe calcular en 7 años. Por otro lado, un cliente que no ha realizado up-selling pero que es propenso a realizarlo y se le aplica la oferta, estará entregando un porcentaje de dinero mayor a la compañía desde antes de su motivación personal. Este periodo de antelación se considera como 1 año en promedio. La diferencia de la rentabilidad con Plan Multimedia y la rentabilidad sin Plan Multimedia se denomina ME y su flujo en el futuro se ve de la siguiente manera.

Figura 65: Flujos futuros de clientes up-seller.



Fuente: Elaboración Propia.

El monto entregado (ME) puede tomar dos valores:

- Cliente SUS up-seller: $ME = \$31.773 - \$19.000 = \$12.773$
- Cliente CC up-seller: $ME = \$18.990 - \$13.000 = \$5.990$

Es necesario traer a valor presente los flujos futuros que reportaran los clientes up-sellers por el adelanto en el cambio de plan. Para ello, se utiliza una tasa de descuento del 10%, ya que el negocio es estable y no posee mayor riesgo.

- $VP \text{ ingresos SUS} = \87.031
- $VP \text{ ingresos CC} = \40.814

Luego de los supuestos establecidos para los beneficios, también se deben establecer supuestos para los costos de captación (recomendado por experto de la compañía) que se aplicará solamente una vez dentro del periodo de estudio.

- $\text{Costo de captación SUS} = \3.832
- $\text{Costo de captación CC} = \1.797

Al momento de acceder a un Plan Multimedia el cliente genera tanto ingresos como egresos. Por lo tanto se debe considerar el costo que tiene para la compañía que el cliente accede a un Plan Multimedia un año antes. Para ello, en la figura 66 se muestran los costos por tener los respectivos planes (considerando que el costo para la compañía es el 60% del Plan).

Figura 66: Costo para la compañía SUS y CC.

Posee Plan Multimedia	SUS	CC
NO	\$ 11.400	\$ 7.800
SI	\$ 19.063	\$ 11.394

Fuente: Elaboración Propia.

Los costos futuros presentan la misma analogía que los ingresos futuros, como se muestra en la figura 67.

Figura 67: Costos futuros de clientes up-seller.



Fuente: Elaboración Propia.

En definitiva el costo extra (CE) que asumirá la compañía puede tomar dos valores:

- Cliente SUS up-seller: $CE = \$19.063 - \$11.400 = \$7.663$
- Cliente CC up-seller: $CE = \$11.394 - \$7.800 = \$3.594$

Del mismo modo como se hizo para los ingresos, se debe traer a valor presente los gastos futuros que reportaran los clientes up-sellers por el adelanto en el cambio de plan.

- $VP \text{ costos SUS} = \52.213
- $VP \text{ costos CC} = \$24.488$

Por otra parte, según el análisis descriptivo y la aplicación de los modelos se puede definir que para los clientes SUS la participación de Planes Multimedia es de 13,98% y para el caso de los clientes CC la participación de Planes Multimedia es de 4,50%. Estos valores ayudan a deducir la posible tasa de captación de cada uno de los tipos de clientes. De esta manera se estima que por la predicción del modelo y el comportamiento creciente de los Planes Multimedia estos últimos años la tasa de captación sean las siguientes.

- $Tasa \text{ de captación SUS} = 6,99\%$
- $Tasa \text{ de captación CC} = 2,25\%$

Finalmente, los ingresos para la compañía serán.

$$\text{Clientes Potenciales} = CP$$

$$\text{Tasa de captación} = TC$$

$$\begin{aligned} \text{Ingresos} = & [CP \text{ SUS} * TC \text{ SUS} * VP \text{ ingresos SUS} - CP \text{ SUS} * TC \text{ SUS} \\ & * \text{Costo de captación SUS} - CP \text{ SUS} * TC \text{ SUS} * VP \text{ costos SUS}] \\ & + [CP \text{ CC} * TC \text{ CC} * VP \text{ ingresos CC} - CP \text{ CC} * TC \text{ CC} \\ & * \text{Costo de captación CC} - CP \text{ CC} * TC \text{ CC} * VP \text{ costos CC}] \end{aligned}$$

$$\Rightarrow \text{Ingresos} = [11753 * 0,069 * 87031 - 11753 * 0,069 * 3832 - 11753 * 0,069 * 52213] + [16898 * 0,0225 * 40814 - 16898 * 0,0225 * 1797 - 16898 * 0,0225 * 24488]$$

$$\text{Ingresos} = \$30.652.312$$

Este resultado corresponde al valor presente de los flujos futuros de la estrategia comercial. Además, se puede observar que solo corresponde al 10% del universo de datos, por lo tanto perfectamente este resultado puede ser extrapolable.

8.7 PRESENTACIÓN DEL CONOCIMIENTO Y CONCLUSIONES

El presente trabajo de título buscó generar ofertas focalizadas para segmentos de clientes propensos a comprar Planes Multimedia. Se llevó a cabo a través de la metodología KDD (Knowledge Discovery for Data Base). Con esta metodología, se buscó descubrir patrones ocultos en grandes cantidades de datos recopiladas. El enfoque del aprendizaje adquirido durante el desarrollo del proyecto se enfocó en dos grandes hitos. Por una parte, el análisis descriptivo y unidimensional de los datos y por otra parte, y quizás la más importante, la aplicación de tres técnicas de minería de datos. Las técnicas utilizadas fueron Árboles de Decisión, Redes Neuronales y Regresiones Logísticas.

Hoy en día, comprender las necesidades de los clientes se ha transformado en el desafío de toda empresa. Pues, en un negocio como la telefonía móvil que se caracteriza por tener un alto grado de competencia, rápidos cambios tecnológicos y una alta rentabilidad, cada día es más importante conocer el comportamiento de los clientes para poder entregar el servicio correcto en el momento correcto. En específico, Entel S.A., no está enfocado en realizar esfuerzos por atraer a nuevos clientes, sino que su objetivo es aumentar el ingreso promedio por usuario. Sin duda que lo anterior está ligado a entregar servicios de valor agregado en conjunto con desarrollo tecnológico de las redes y Smartphone que apoyan a la oferta de la compañía.

El valor agregado, de este trabajo de título, se encuentra en la generación de reglas para predecir la propensión de compra de clientes con alta probabilidad a acceder a Planes Multimedia. Por otro lado ayuda a entender el comportamiento de clientes que ya han realizado up-selling y por lo demás aprender de ellos para identificar posible up-seller de la base de clientes con planes antiguos. Si bien, el análisis descriptivo detalla cada una de las tendencias de las variables relevantes y los modelos predicen con alta probabilidad el up-selling, lo que está por detrás es una metodología que absorbe información relevante, cualitativa y cuantitativa, para la gestión de la compañía.

Los resultados por si solos ya son de gran ayuda para las distintas áreas de la empresa, pero además, sirven como paso previo para otros estudios posteriores. Por ejemplo, analizar la presión del entorno de los clientes por acceder a ciertos servicios, indagar en el comportamiento detallado con la red de contacto de cada cliente, analizar quienes son clientes líderes en los servicios que entrega la compañía y quiénes son los seguidores, y por último sirve para levantar requerimientos de variables que son relevantes y no se ha tenido una buena gestión dentro de la empresa.

En particular, la generación de ofertas focalizadas a clientes de baja vinculación, a la compañía, se pudo lograr. Por una parte, de manera más descriptiva, se pudo identificar que los clientes que han realizado up-selling en su mayoría son hombres, entre 21 y 40 años, de sector económico ABC1 o C2, solteros, con estudios Universitarios, de actividad profesional Ingenieros, Minería, Salud, Docencia y Administrativos, con uno a cuatro años de antigüedad en la compañía. Del mismo modo, se puede decir que son clientes que han ido aumentando su facturación en los últimos periodos hasta traducir este comportamiento en un salto drástico en su facturación al momento de realizar el up-selling. Para el tráfico de minutos y el tráfico de mensajes existe una similitud con lo anterior. Su tráfico aumenta con el transcurso del tiempo, sin una gran diferencia con los clientes que no han realizado up-selling, pero al momento de realizar el cambio de plan su tráfico aumenta considerablemente, lo que reporta directamente en mayores ingresos para la compañía. Y por último, en este sentido, la decisión de acceder a un Smartphone está ligada al up-grade a Planes Multimedia.

Por otra parte, desde una mirada más analítica, la generación de distintas técnicas de minería de datos, ayudó a tener un contraste entre cada una de ellas y poder elegir el mejor modelo con una predicción de 83,39%. De esta manera los patrones de comportamiento ayudaron a generar ofertas comerciales a aquellos clientes con consumo elevado de descarga de datos de internet, que no tengan problemas financieros con la compañía, que trafiquen mensajes de texto o mensajes multimedia, que además sean menor de 46 años, que su facturación mensual sea superior a los \$15.000 en promedio, que su tasa de crecimiento de minutos traficado sea creciente mes a mes, que este recibiendo más mensajes multimedia en el último tiempo y que este contratando bolsas de distintos servicios para suplir la falta de servicios en su plan.

Lamentablemente, no todos las técnicas empleadas entregaron buenos resultados. Por lo tanto, se vio limitada, en cierta parte, la generación de ofertas comerciales complementarias entre las distintas técnicas, además de poder haber comparado si las variables explicativas coincidían en los modelos. La posible explicación es la baja población de los datos en variables importantes dentro del análisis del comportamiento, como el sector económico, la profesión, usuario del teléfono, estado civil, entre otras que al realizar un análisis descriptivo detallado y no tan formal aparecían como candidatos para la realización de up-selling. Por lo tanto se deja como acciones de mejora recopilar de mejor manera los datos de los clientes y por sobre todo llevar la data histórica mensual de las cuentas de arriendo de los equipos terminales. Ya que Smartphone y Plan Multimedia van de la mano a la entrega de servicios de valor agregado.

9. BIBLIOGRAFÍA Y FUENTES DE INFORMACIÓN

A continuación se presentan las fuentes de información consultadas para la realización de este trabajo de investigación.

1. **HAN, JIAWES y KAMBER, MICHELINE.** 2001. Data Mining: Concepts and Techniques. Segunda Edición.
2. **BIELZA, CONCHA y LARRAÑAGA, PEDRO.** [s.a.]. Árboles de Clasificación, Aprendizaje Automático. Universidad Politécnica de Madrid, Facultad de Informática, Departamento de Inteligencia Artificial.
3. **MARTÍNEZ MUÑOZ, GONZALO.** 2006. "Clasificación Mediante Conjuntos". Tesis Doctoral. España, Universidad Autónoma de Madrid, Departamento de Ingeniería Informática.
4. **EDWIN D., ONGLEY y DORNER, SARAH.** [s.a.]. Sistema de Apoyo para la toma de Decisiones para el Pronóstico de la Contaminación del Agua en la Agricultura". Canadá, University of Guelph, Departament of Computing and Information Science.
5. **LÓPEZ TAKEYAS, BRUNO.** 2005. Inteligencia Artificial. Tamaulipas, Instituto Tecnológico Nuevo Laredo.
6. **VICENTE TAVERA, SANTIAGO y RAMIREZ, GUILLERMO.** 1994. Un Análisis de Segmentación en el Estudio del Paro en Salamanca. España, Universidad de Salamanca, Departamento de Estadística y Matemáticas Aplicadas.
7. **RULEQUEST RESARCH, DATA MINING TOOLS.** [s.a.]. Es el modelo C5.0 mejor que el C4.5. [en línea]. [Consulta: 22.agosto 2011] <<http://www.rulequest.com/see5-comparison.html>>
8. **BASUALDO, MARTA y RUIZ, CARLOS.** [s.a.]. Redes Neuronales: Conceptos Básicos y Aplicaciones. Cátedra: Informática Aplicada a la Ingeniería de Procesos. Argentina, Universidad Tecnológica Nacional, Departamento de Ingeniería Química.
9. **ROJO ABUÍN, JOSÉ MANUEL.** 2007. Regresión con Variable Dependiente Cualitativa. España, Instituto de Economía y Geografía.
10. **PALMA, MATILDE y TRUJILLANO, JAVIER.** [s.a.]. Approach to the methodology of classification and regression trees. [en línea]. [Consulta: 22.agosto 2011] <http://scielo.isciii.es/scielo.php?pid=S021391112008000100013&script=sci_arttext>.
11. **RUIZ REYES, MARCELO EDUARDO.** 2004. Identificación de clientes en una empresa de telecomunicaciones con alta probabilidad de realizar venta cruzada

utilizando técnicas de data mining. Memoria (Ingeniería Civil Industrial). Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

12. **CARVAJAL CARTAGENA, JUAN PABLO.** 2002. Desarrollo de un modelo de propensión de compra para una empresa de comunicaciones. Memoria (Ingeniería Civil Industrial). Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

13. **PRIDAL HERNANDEZ, TOMAS JAROMIR.** 2008. Propuesta de una segmentación de clientes en una industria de servicios B2B: Caso Telefónica empresas. Memoria (Ingeniería Civil Industrial). Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

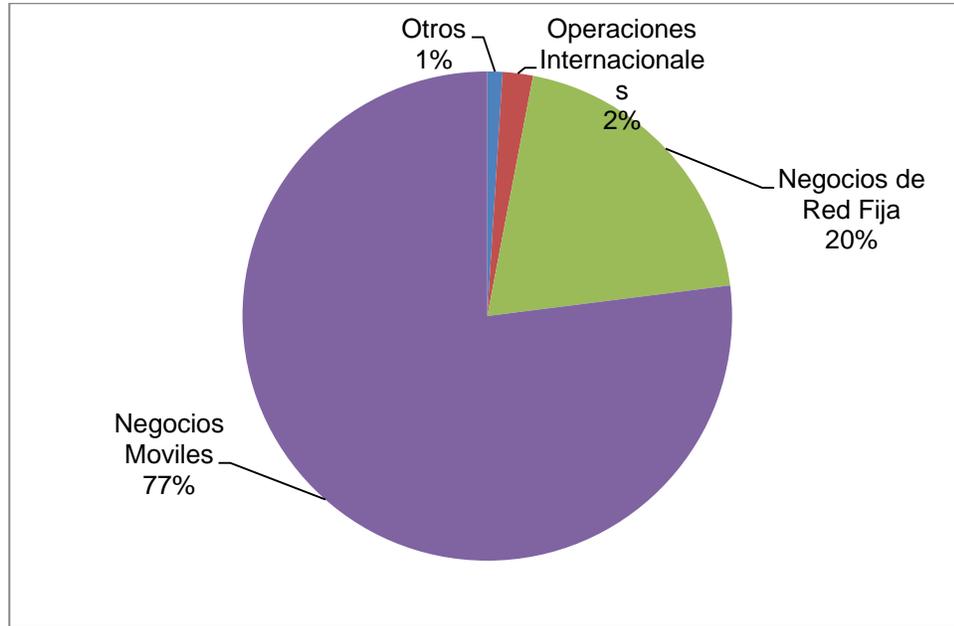
14. **PARRAGUEZ VERDUGO, MANUEL ALEJANDRO.** 2010. Generación de ofertas focalizadas para clientes de baja vinculación en una institución bancaria. Memoria (Ingeniería Civil Industrial). Chile, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.

15. **HAND y MANNILA.** 2001. Principles of Data Mining. Tercera Edición.

16. **CHAPMAN, PETE y CLINTON, JULIAN.** 2000. CRISP-DM 1.0, Step by Step data mining guide.

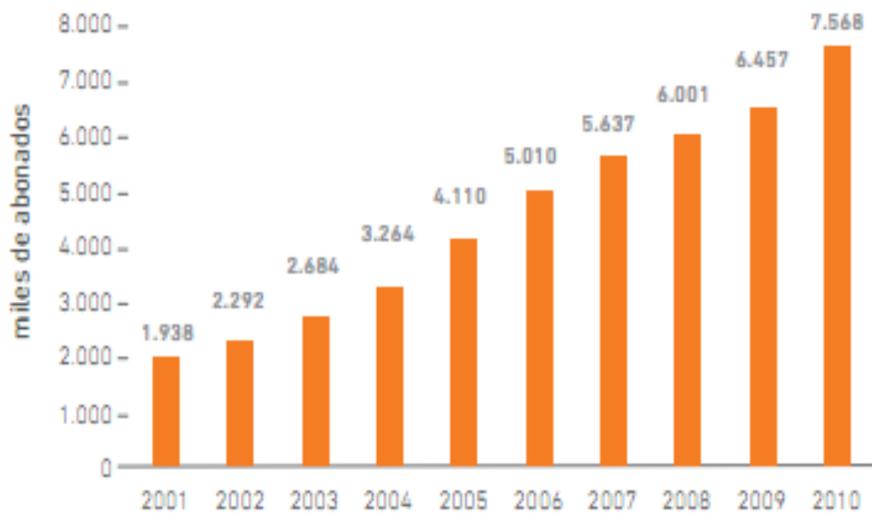
10. ANEXOS

Anexo 1: Distribución de Ingresos Consolidados por Negocio.



Fuente: Memoria Anual Entel S.A., 2010.

Anexo 2: Evolución Base de Abonados Negocio Móvil.



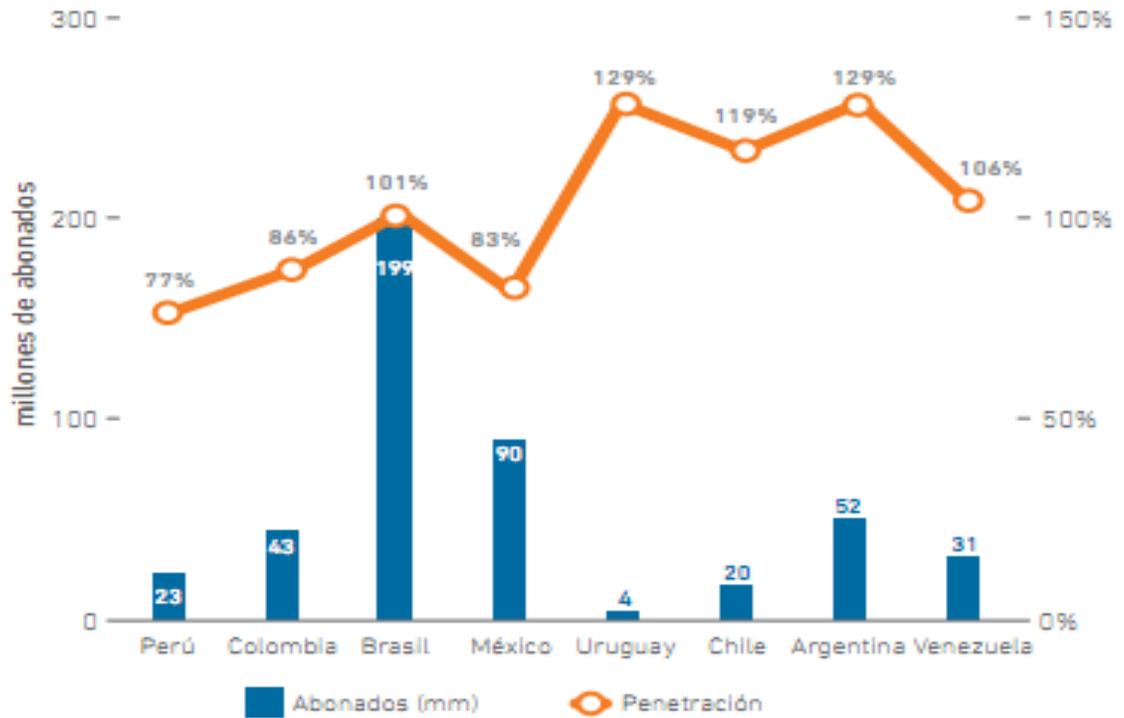
Fuente: Memoria Anual Entel S.A., 2010.

Anexo 3: Evolución Penetración y Abonados en la Telefonía Móvil en el país.

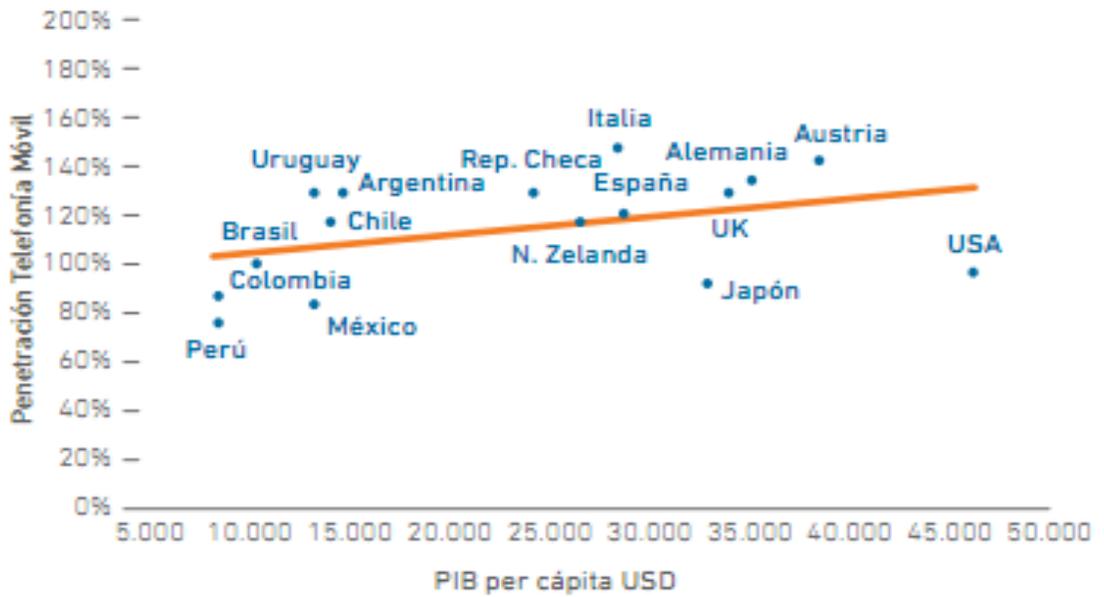


Fuente: Memoria Anual Entel S.A., 2010.

Anexo 4: Índices de Penetración de Telefonía Móvil en Latinoamérica.

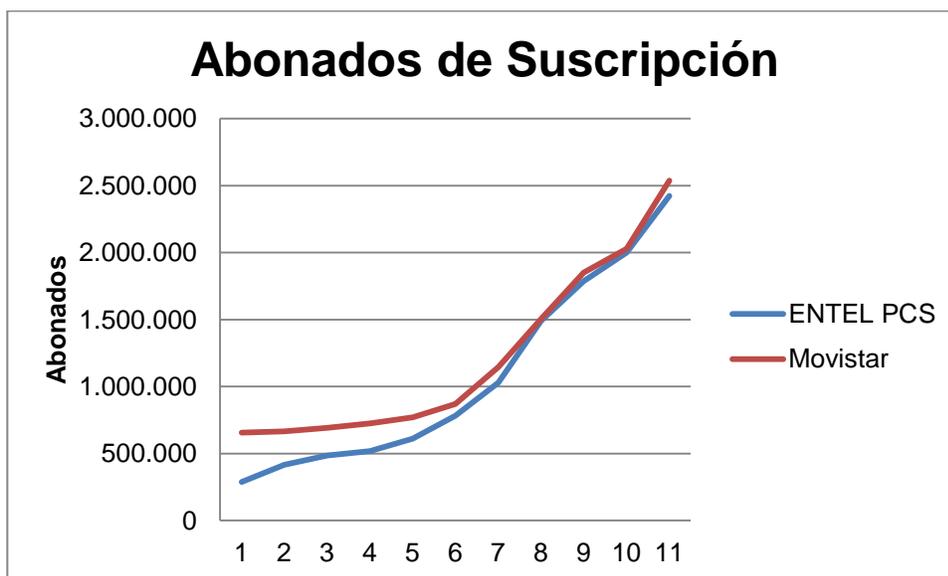
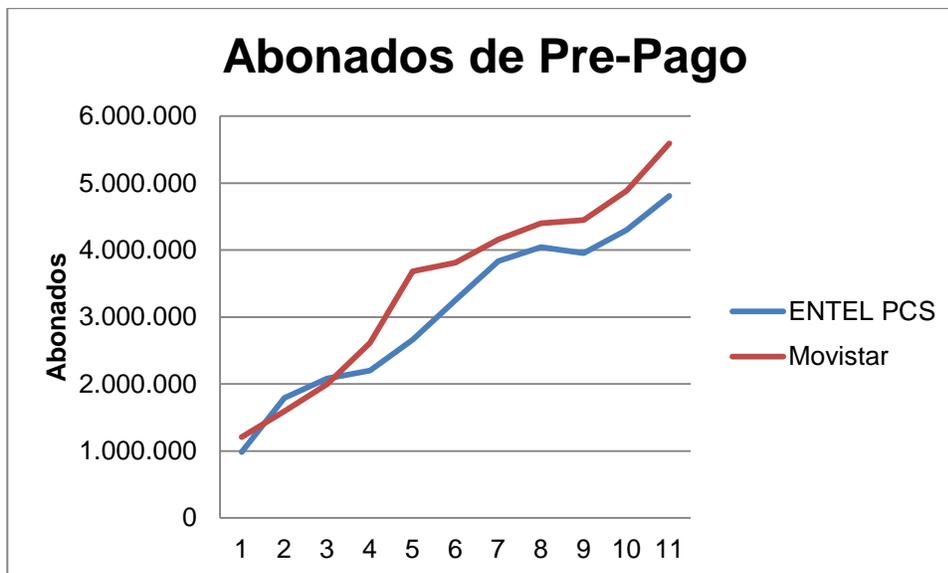


Penetración Telefonía Móvil versus PIB per Cápita.



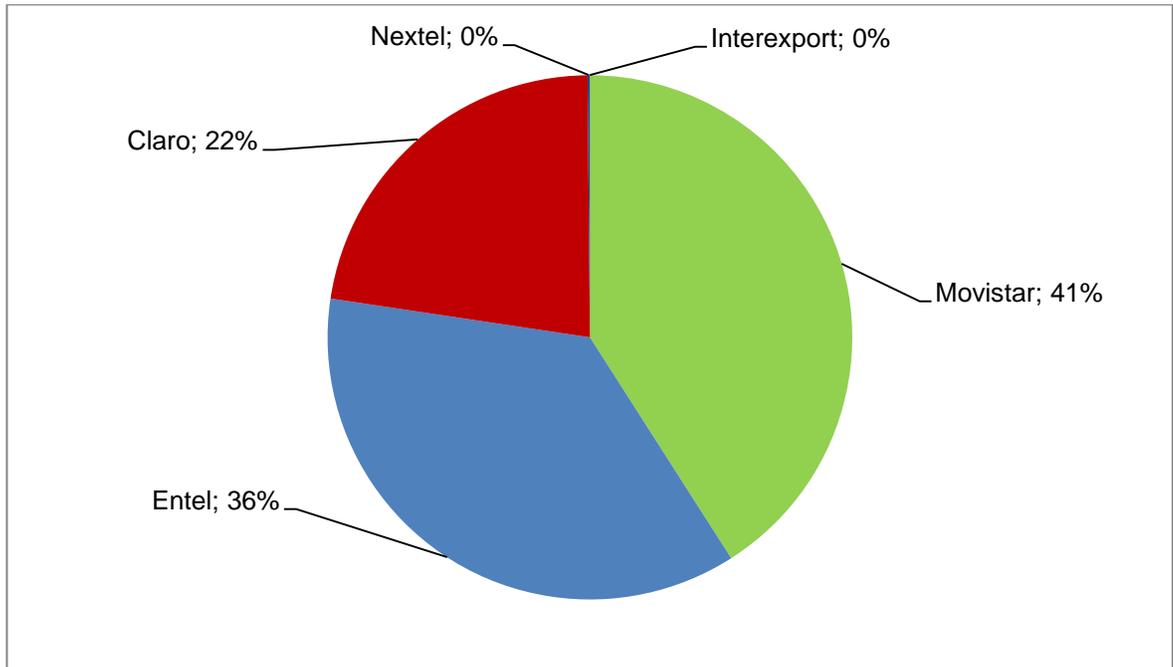
Fuente: Memoria Anual Entel S.A., 2010.

Anexo 5: Participación de los principales competidores en Telefonía Móvil medida por el número de abonados.

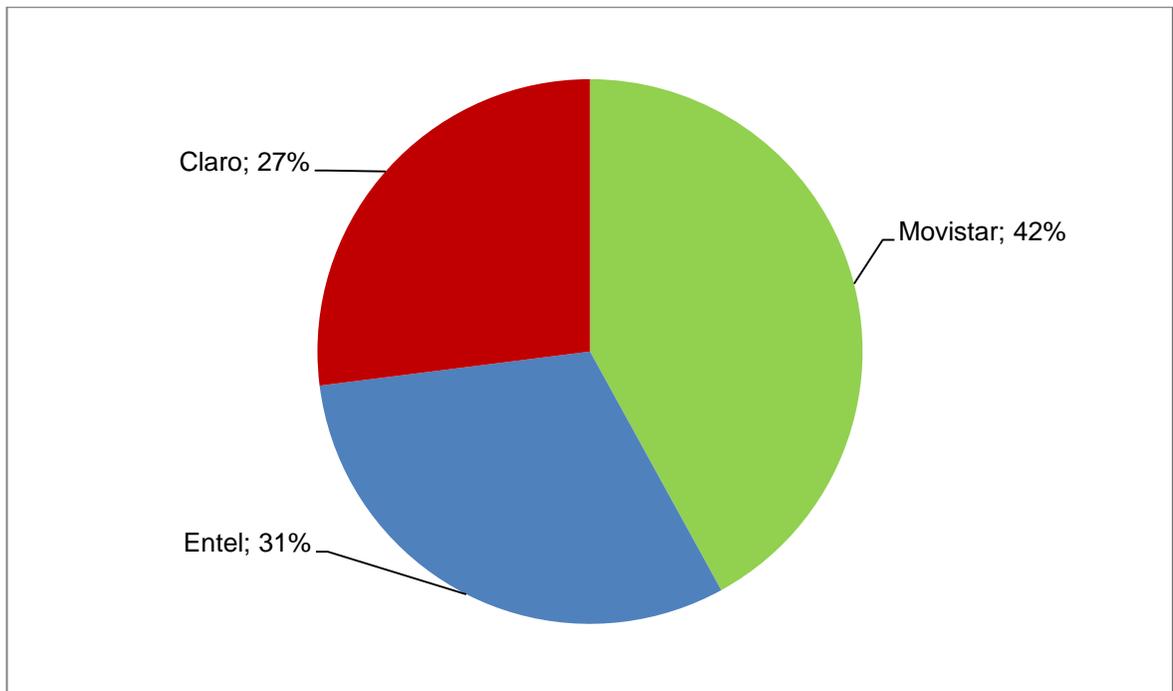


Fuente: Estadísticas Telefonía Móvil SUBTEL, 2010.

Anexo 6: Participación de las empresas de Telefonía Móvil medida por el número de abonados.

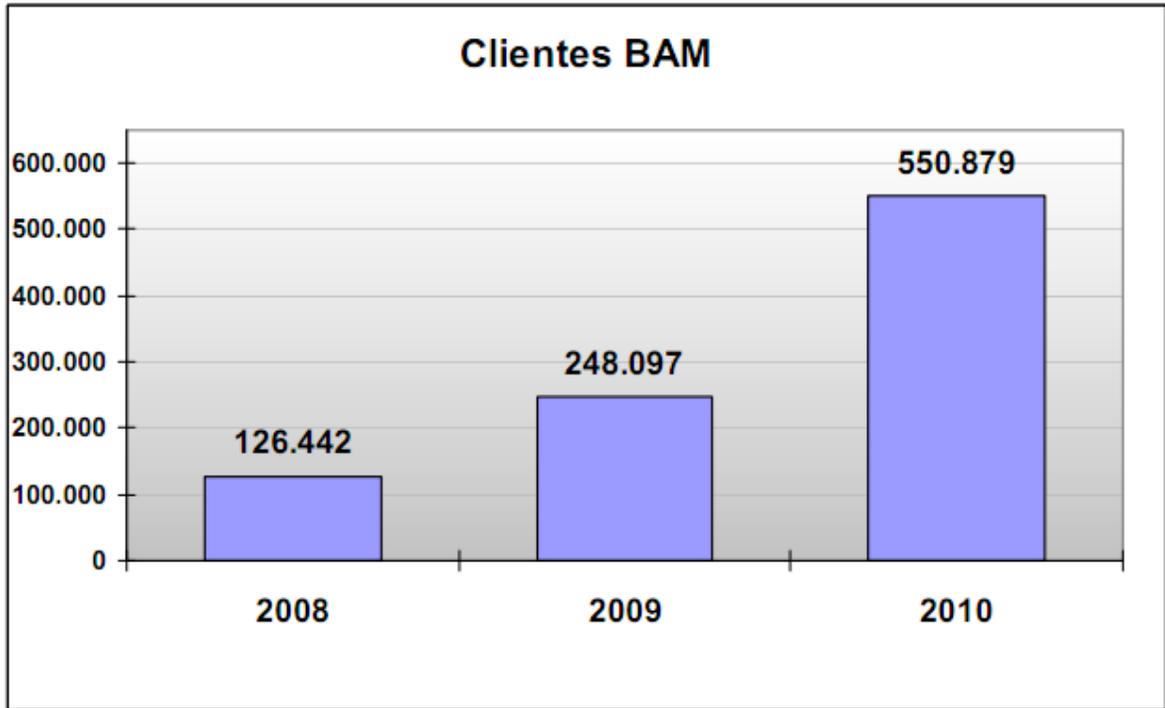


Participación de las empresas de Internet Móvil medida por el número de abonados.



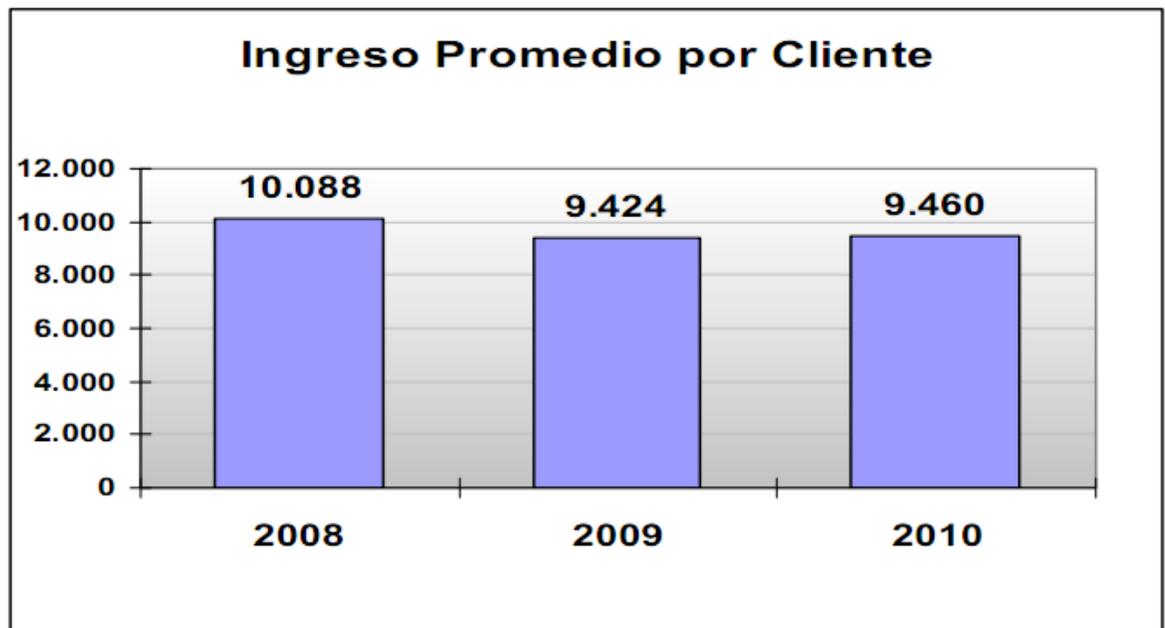
Fuente: Estadísticas Telefonía Móvil SUBTEL, 2010.

Anexo 7: Evolución Base de Clientes BAM.



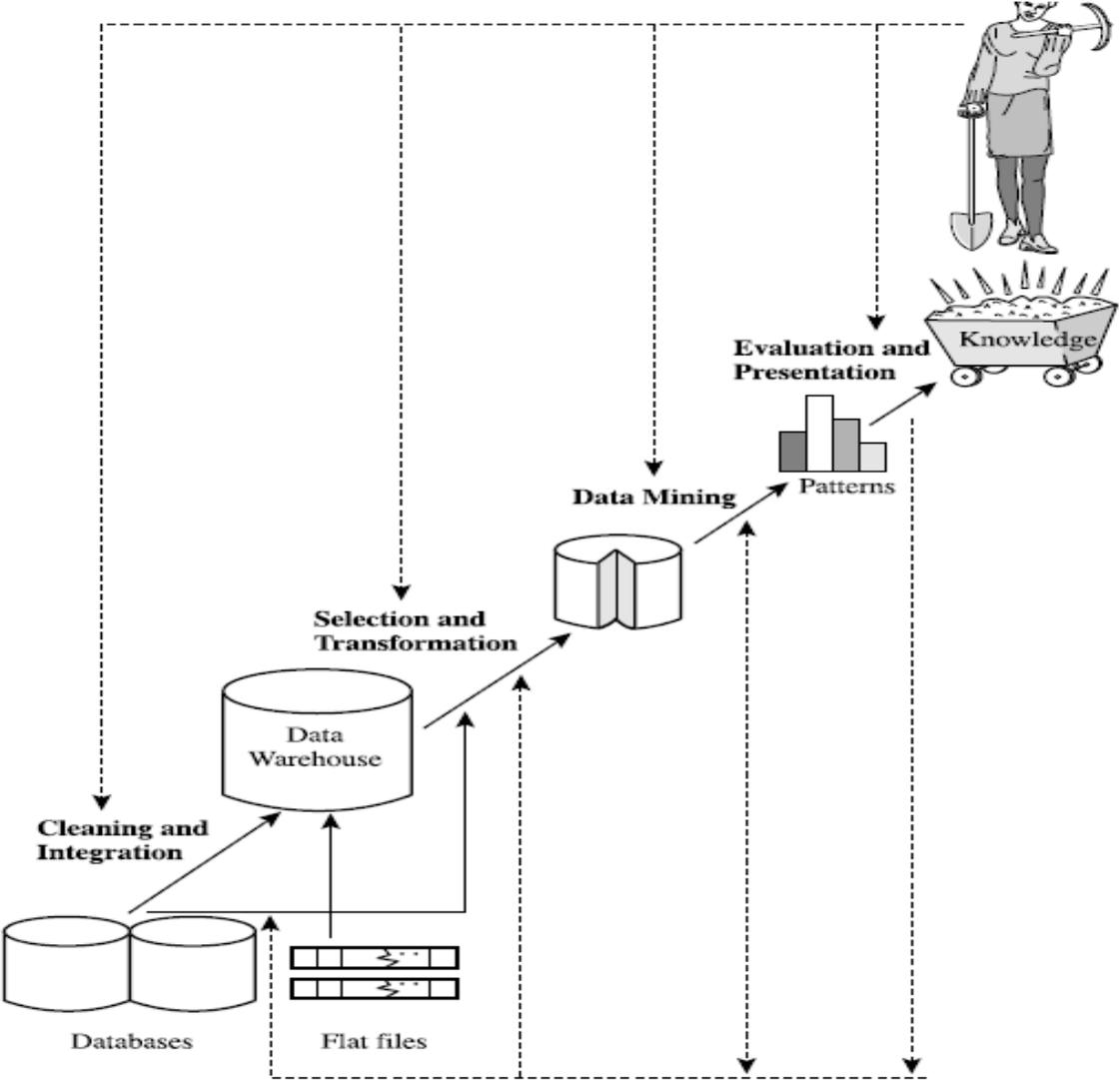
Fuente: Memoria Anual Entel S.A., 2010.

Anexo 8: Evolución Ingreso Promedio por Cliente (ARPU).



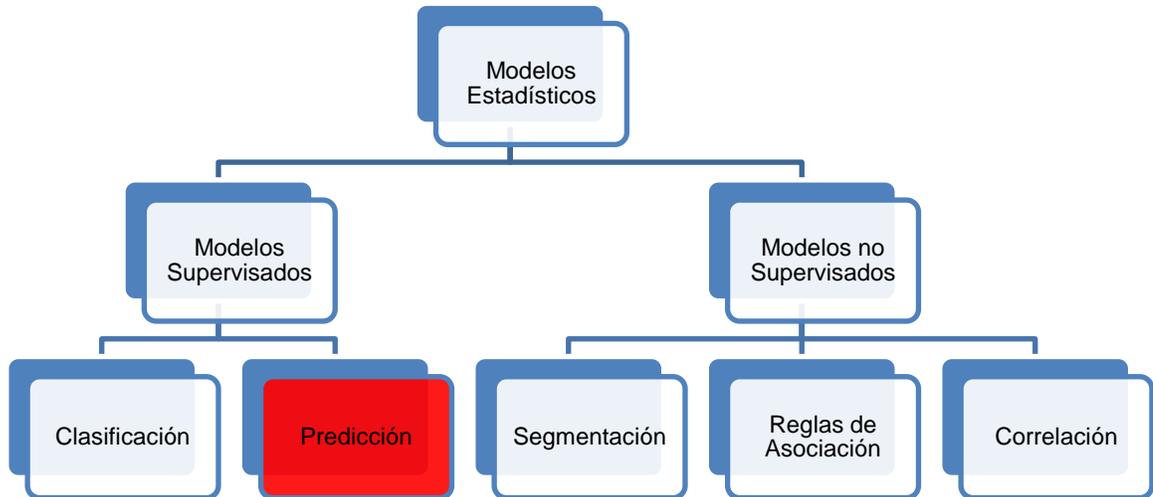
Fuente: Memoria Anual Entel S.A., 2010.

Anexo 9: Metodología KDD.

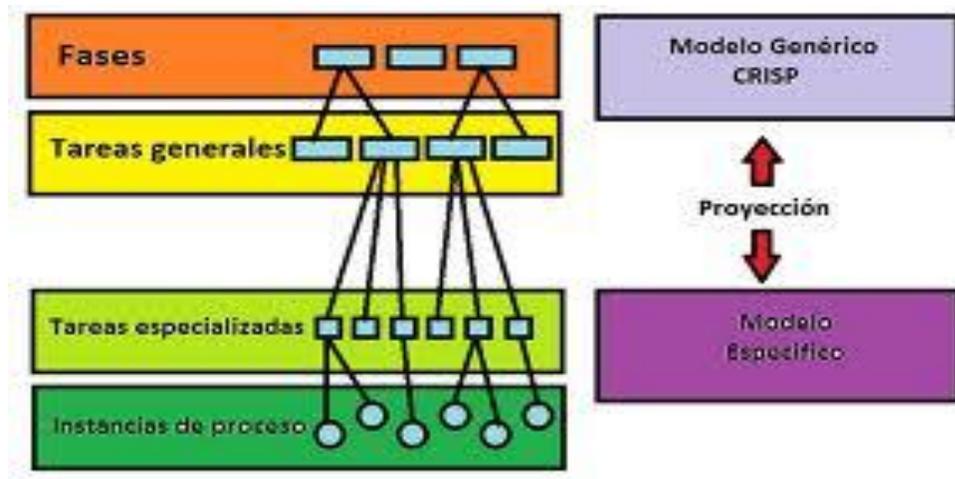


Fuente: Data Mining, Concepts and Techniques.

Anexo 10: Modelos Supervisados y Modelos no Supervisados

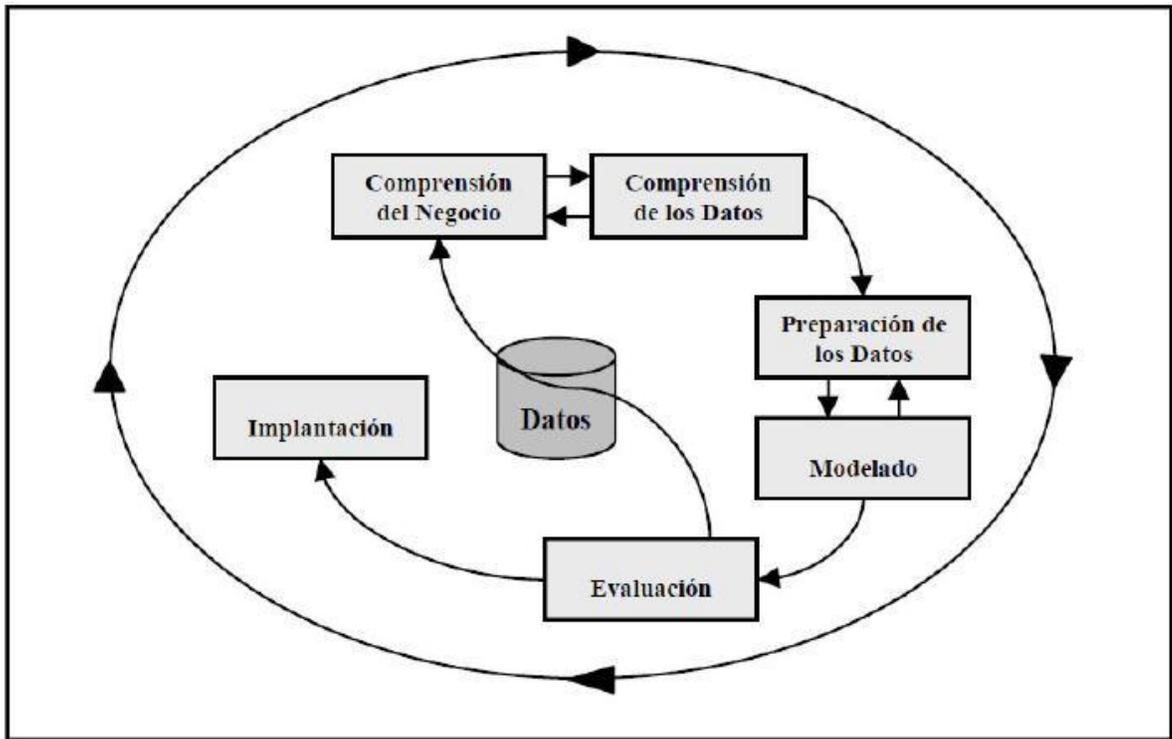


Anexo 11: Modelo de proceso jerárquico CRISP DM.



Fuente: CRISP-DM 1.0. Step by step data mining guide.

Anexo 12: Modelo CRISP DM.



Fuente: CRISP-DM 1.0. Step by step data mining guide.

Anexo 13: Descripción de las tablas de datos.

BSCS		
Descripción	Campos	Observación
Datos transaccionales e identificadores del cliente	Número de contrato	Información mensual
	Número móvil	
	Rut del Cliente	
	Descripción del plan	
	Nombre del Cliente	
	Dirección del cliente	
	Cuidad del cliente	
	Comuna del cliente	
	Número de facturación	
	Estado actual del móvil	
	Fecha ingreso a la compañía	
	Grupo del cliente	
	Fecha último cambio de estado	

	Motivo cambio de estado	
	Plan original	

BUIC		
Descripción	Campos	Observación
Datos demográficos y geográficos	Número móvil	Información correspondiente a Septiembre del 2011
	Descripción del plan	
	Fecha ingreso a la compañía	
	Rut del cliente	
	Usuario o dueño	
	Nombre del cliente	
	Género	
	Fecha de nacimiento	
	Estado civil	
	Dirección del cliente	
	Email	
	Nivel de estudio	
	Actividad	
	Profesión	
	Hijos	
Fecha última modificación		
Cuidad		
Comuna		
Sector económico		

TERMINALES		
Descripción	Campos	Observación
Datos del equipo del cliente	Número móvil	Información mensual
	Marca del equipo	
	Modelo del equipo	
	Tecnología del equipo	

TRÁFICO DE MINUTOS		
Descripción	Campos	Observación
Datos transaccionales de tráfico de minutos	Número móvil	Información mensual
	Minutos emitidos	
	Minutos recibidos	

TRÁFICO DE MENSAJERÍA		
Descripción	Campos	Observación
Datos transaccionales de tráfico de mensajes	Número móvil	Información mensual
	Mensaje de texto emitidos	
	Mensaje de texto recibidos	
	Mensajes multimedia emitidos	
	Mensajes multimedia recibidos	

TRÁFICO DE DATOS		
Descripción	Campos	Observación
Datos transaccionales de tráfico de datos	Número móvil	Información mensual
	kbytes mensajes multimedia	
	kbytes internet móvil	
	kbytes internet banda ancha	

FACTURACIÓN		
Descripción	Campos	Observación
Datos de facturación	Número móvil	Información mensual
	Descripción del plan	
	Concepto de facturación	
	Unidad del servicio traficado	
	Cantidad de tráfico	
	Monto facturado	

RECARGAS		
Descripción	Campos	Observación
Datos de recargas	Número móvil	Información mensual
	Cantidad de recargas	
	Monto de la recarga	
	Bono de recarga	

BOLSAS		
Descripción	Campos	Observación

Datos transaccionales de bolsas de servicios	Número móvil	Información mensual
	Tipo de bolsa	
	Monto de la bolsa	
	Cantidad de servicio	

ACCIONES DE COBRANZA		
Descripción	Campos	Observación
Datos de riesgo (acciones de cobranza)	Número de contrato	Información mensual
	Acción de cobranza	
	Rut del cliente	

ATRASO PROMEDIO		
Descripción	Campos	Observación
Datos de riesgo (atraso promedio)	Rut del cliente	Información acumulada de los últimos 6 meses
	Suma de atrasos	
	Atraso promedio	

ACOC		
Descripción	Campos	Observación
Datos cuotas de arriendo equipo terminales	Rut del cliente	Información acumulada histórica
	Numero de contrato	
	Estado de pago	
	Cuotas	
	Cuotas pagadas	
	Valor de la cuota	
	Valor de la opción	

Anexo 14: Pesos de la Regresión Logística para clientes SUS

Attribute	Weight
CF_(t-1)	-0.191
Min_(t-1)	0.183
anos_entel	0.041
edad	0.365
numero_lineas	-0.223
acoc_faltantes_11'	-0.025
min_emi_(t-1)	-0.199
min_emi_(t-2)	-0.346
min_emi_(t-3)	-0.114
min_emi_(t-4)	-0.172
min_emi_(t-5)	0.271
min_emi_(t-6)	-0.282
prom_min_emi_1-3	0.786
prom_min_emi_4-6	0.143
dif_min_emi_1-3_4-6	0.022
r_crec_min_emi_1-3_4-6	-0.001
%_min_emi_plan	0.486
%_min_emi_plan_sinexceso	-0.090
%_min_emi_plan_exceso	-0.098
categoria_exceso_min_emi	0.240
%_min_emi_historico	-0.001
%_min_emi_plan_sinexceso_historico	0.071
%_min_emi_plan_exceso_historico	0.143
categoria_exceso_min_emi_historico	-0.164
r_crec_min_emi_1	0.213
r_crec_min_emi_2	-0.077
r_crec_min_emi_3	-0.100

Attribute	Weight
r_crec_min_emi_4	0.085
r_crec_min_emi_5	-0.054
r_crec_min_emi_prom	-0.010
min_recep_(t-1)	-0.023
min_recep_(t-2)	0.394
min_recep_(t-3)	0.093
min_recep_(t-4)	0.001
min_recep_(t-5)	0.178
min_recep_(t-6)	-0.255
prom_min_recep_1-3	-0.296
prom_min_recep_4-6	-0.308
dif_min_recep_1-3_4-6	-0.198
r_crec_min_recep_1-3_4-6	-0.061
%_min_recep_historico	-0.061
%_min_recep_plan_sinexceso_historico	-0.028
%_min_recep_plan_exceso_historico	0.139
categoria_exceso_min_recep_historico	0.110
r_crec_min_recep_1	-0.050
r_crec_min_recep_2	-0.048
r_crec_min_recep_3	-0.500
r_crec_min_recep_4	-0.164
r_crec_min_recep_5	0.006
r_crec_min_recep_prom	0.421
sms_mms_emi_(t-1)	0.199
sms_mms_emi_(t-2)	0.624
sms_mms_emi_(t-3)	0.296
sms_mms_emi_(t-4)	0.400
sms_mms_emi_(t-5)	0.356

Attribute	Weight
sms_mms_emi_(t-6)	0.189
prom_sms_mms_emi_1-3	-1.169
prom_sms_mms_emi_4-6	-0.748
dif_sms_mms_emi_1-3_4-6	0.018
r_crec_sms_mms_emi_1-3_4-6	1.108
%_sms_mms_emi_historico	0.142
%_sms_mms_emi_plan_sinexceso_historico	-0.189
%_sms_mms_emi_plan_exceso_historico	-0.896
categoria_exceso_sms_mms_emi_historico	-0.152
r_crec_sms_mms_emi_1	-0.163
r_crec_sms_mms_emi_2	-0.042
r_crec_sms_mms_emi_3	-0.474
r_crec_sms_mms_emi_4	0.028
r_crec_sms_mms_emi_5	0.005
r_crec_sms_mms_emi_prom	0.126
sms_mms_recep_(t-1)	-0.580
sms_mms_recep_(t-2)	-0.651
sms_mms_recep_(t-3)	-0.502
sms_mms_recep_(t-4)	0.694
sms_mms_recep_(t-5)	0.598
sms_mms_recep_(t-6)	0.325
prom_sms_mms_recep_1-3	0.806
prom_sms_mms_recep_4-6	-0.620
dif_sms_mms_recep_1-3_4-6	0.801
r_crec_sms_mms_recep_1-3_4-6	0.803
%_sms_mms_recep_historico	-0.046
%_sms_mms_recep_plan_sinexceso_historico	-0.023
%_sms_mms_recep_plan_exceso_historico	-1.243

Attribute	Weight
%_sms_mms_recep_plan_exceso_historico	-1.243
categoria_exceso_sms_mms_recep_historico	-0.120
r_crec_sms_mms_recep_1	0.060
r_crec_sms_mms_recep_2	0.199
r_crec_sms_mms_recep_3	0.014
r_crec_sms_mms_recep_4	-0.229
r_crec_sms_mms_recep_5	-0.384
r_crec_sms_mms_recep_prom	0.081
ingresos_(t-1)	-0.250
ingresos_(t-2)	0.081
ingresos_(t-3)	-0.369
ingresos_(t-4)	-0.093
ingresos_(t-5)	-0.201
ingresos_(t-6)	-0.362
prom_ingresos_1-3	0.286
prom_ingresos_4-6	0.352
dif_ingresos_1-3_4-6	0.049
r_crec_ingresos_1-3_4-6	0.866
%_ingresos_plan	-1.946
%_ingresos_plan_sinexceso	0.004
%_ingresos_plan_exceso	1.686
categoria_exceso_ingresos	-0.220
%_ingresos_historico	-0.312
%_ingresos_plan_sinexceso_historico	-0.053
%_ingresos_plan_exceso_historico	-0.172
categoria_exceso_ingresos_historico	-0.294
r_crec_ingresos_1	0.021
r_crec_ingresos_2	0.126

Attribute	Weight
r_crec_ingresos_3	0.254
r_crec_ingresos_4	0.005
r_crec_ingresos_5	0.029
r_crec_ingresos_prom	-0.001
q_recargas_(t-1)	-0.205
q_recargas_(t-2)	0.133
q_recargas_(t-3)	0.021
q_recargas_(t-4)	0.165
q_recargas_(t-5)	0.313
q_recargas_(t-6)	0.145
prom_q_recargas_1-3	0.075
prom_q_recargas_4-6	-0.227
dif_q_recargas_1-3_4-6	0.215
r_crec_q_recargas_1-3_4-6	0.139
%_q_recargas_historico	0.058
%_q_recargas_plan_sinexceso_historico	-0.098
%_q_recargas_plan_exceso_historico	0.345
categoria_exceso_q_recargas_historico	-0.420
r_crec_q_recargas_1	0.048
r_crec_q_recargas_2	0.005
r_crec_q_recargas_3	-0.003
r_crec_q_recargas_4	0.022
r_crec_q_recargas_5	-0.071
r_crec_q_recargas_prom	-0.111
n_acc_cobranza_(t-1)	2.046
n_acc_cobranza_(t-2)	-0.231
n_acc_cobranza_(t-3)	-0.301
n_acc_cobranza_(t-4)	-0.205

Attribute	Weight
n_acc_cobranza_(t-5)	0.255
n_acc_cobranza_(t-6)	0.923
prom_n_acc_cobranza_1-3	-0.605
prom_n_acc_cobranza_4-6	-0.213
dif_n_acc_cobranza_1-3_4-6	0.175
r_crec_n_acc_cobranza_1-3_4-6	0.370
%_n_acc_cobranza_historico	1.564
%_n_acc_cobranza_plan_sinexceso_historico	-1.179
%_n_acc_cobranza_plan_exceso_historico	-1.353
categoria_exceso_n_acc_cobranza_historico	0.032
r_crec_n_acc_cobranza_1	-0.492
r_crec_n_acc_cobranza_2	-0.332
r_crec_n_acc_cobranza_3	-0.382
r_crec_n_acc_cobranza_4	-0.027
r_crec_n_acc_cobranza_5	-0.087
r_crec_n_acc_cobranza_prom	0.674
suma_atraso	-0.001
atraso_promedio	-0.045
n_bolsas_(t-1)	-0.024
n_bolsas_(t-2)	-0.004
n_bolsas_(t-3)	0.050
n_bolsas_(t-4)	0.382
n_bolsas_(t-5)	0.067
n_bolsas_(t-6)	0.919
prom_n_bolsas_1-3	-0.053
prom_n_bolsas_4-6	-1.514
dif_n_bolsas_1-3_4-6	-0.198
r_crec_n_bolsas_1-3_4-6	-0.057

Attribute	Weight
%_n_bolsas_historico	0.060
%_n_bolsas_plan_sinexceso_historico	-0.424
%_n_bolsas_plan_exceso_historico	-0.090
categoria_exceso_n_bolsas_historico	-0.149
r_crec_n_bolsas_1	0.450
r_crec_n_bolsas_2	0.526
r_crec_n_bolsas_3	0.710
r_crec_n_bolsas_4	0.428
r_crec_n_bolsas_5	0.569
r_crec_n_bolsas_prom	-0.835
kb_gprs_(t-1)	1.003
kb_gprs_(t-2)	0.641
kb_gprs_(t-3)	0.433
kb_gprs_(t-4)	-0.662
kb_gprs_(t-5)	1.000
kb_gprs_(t-6)	-0.589
prom_kb_gprs_1-3	4.610
prom_kb_gprs_4-6	-4.305
dif_kb_gprs_1-3_4-6	-5.434
r_crec_kb_gprs_1-3_4-6	-0.806
consumo_kb_gprs	0.723

Anexo 15: Pesos de la Regresión Logística para clientes CC

Attribute	Weight
CF_(t-1)	-0.493
Min_(t-1)	-0.042
anos_entel	-0.059
edad	0.445
numero_lineas	0.125
acoc_faltantes_11'	-0.035
min_emi_(t-1)	0.006
min_emi_(t-2)	-0.070
min_emi_(t-3)	0.546
min_emi_(t-4)	-0.791
min_emi_(t-5)	0.025
min_emi_(t-6)	-0.107
prom_min_emi_1-3	0.283
prom_min_emi_4-6	0.571
dif_min_emi_1-3_4-6	-0.210
r_crec_min_emi_1-3_4-6	0.176
%_min_emi_plan	0.579
%_min_emi_plan_sinexceso	0.006
%_min_emi_plan_exceso	-0.597
categoria_exceso_min_emi	-0.239
%_min_emi_historico	0.176
%_min_emi_plan_sinexceso_historico	0.063
%_min_emi_plan_exceso_historico	0.421
categoria_exceso_min_emi_historico	-0.101
r_crec_min_emi_1	-0.037
r_crec_min_emi_2	0.087
r_crec_min_emi_3	-0.158

Attribute	Weight
r_crec_min_emi_4	-0.201
r_crec_min_emi_5	0.044
r_crec_min_emi_prom	0.133
min_recep_(t-1)	-0.358
min_recep_(t-2)	-0.470
min_recep_(t-3)	-0.645
min_recep_(t-4)	0.333
min_recep_(t-5)	0.531
min_recep_(t-6)	0.115
prom_min_recep_1-3	0.214
prom_min_recep_4-6	-0.182
dif_min_recep_1-3_4-6	0.547
r_crec_min_recep_1-3_4-6	0.162
%_min_recep_historico	0.162
%_min_recep_plan_sinexceso_historico	-0.034
%_min_recep_plan_exceso_historico	-0.430
categoria_exceso_min_recep_historico	0.110
r_crec_min_recep_1	-0.051
r_crec_min_recep_2	-0.360
r_crec_min_recep_3	0.080
r_crec_min_recep_4	-0.406
r_crec_min_recep_5	0.139
r_crec_min_recep_prom	0.384
sms_mms_emi_(t-1)	0.842
sms_mms_emi_(t-2)	1.277
sms_mms_emi_(t-3)	0.982
sms_mms_emi_(t-4)	-0.100

Attribute	Weight
sms_mms_emi_(t-5)	-0.164
sms_mms_emi_(t-6)	0.158
prom_sms_mms_emi_1-3	-1.372
prom_sms_mms_emi_4-6	-1.122
dif_sms_mms_emi_1-3_4-6	-0.974
r_crec_sms_mms_emi_1-3_4-6	1.220
%_sms_mms_emi_historico	-0.018
%_sms_mms_emi_plan_sinexceso_historico	-0.315
%_sms_mms_emi_plan_exceso_historico	-1.099
categoria_exceso_sms_mms_emi_historico	0.142
r_crec_sms_mms_emi_1	-0.206
r_crec_sms_mms_emi_2	-0.260
r_crec_sms_mms_emi_3	-0.158
r_crec_sms_mms_emi_4	-0.018
r_crec_sms_mms_emi_5	-0.120
r_crec_sms_mms_emi_prom	0.329
sms_mms_recep_(t-1)	-0.219
sms_mms_recep_(t-2)	-0.294
sms_mms_recep_(t-3)	0.251
sms_mms_recep_(t-4)	0.249
sms_mms_recep_(t-5)	0.581
sms_mms_recep_(t-6)	0.250
prom_sms_mms_recep_1-3	0.980
prom_sms_mms_recep_4-6	-1.454
dif_sms_mms_recep_1-3_4-6	-0.598
r_crec_sms_mms_recep_1-3_4-6	-0.047
%_sms_mms_recep_historico	0.279

Attribute	Weight
%_sms_mms_recep_plan_sinexceso_his	0.090
%_sms_mms_recep_plan_exceso_histor	-0.482
categoria_exceso_sms_mms_recep_hist	-0.050
r_crec_sms_mms_recep_1	0.039
r_crec_sms_mms_recep_2	0.279
r_crec_sms_mms_recep_3	-0.206
r_crec_sms_mms_recep_4	0.276
r_crec_sms_mms_recep_5	0.070
r_crec_sms_mms_recep_prom	-0.229
ingresos_(t-1)	0.079
ingresos_(t-2)	-0.061
ingresos_(t-3)	0.037
ingresos_(t-4)	0.550
ingresos_(t-5)	-0.049
ingresos_(t-6)	-0.604
prom_ingresos_1-3	-0.393
prom_ingresos_4-6	-0.785
dif_ingresos_1-3_4-6	-0.473
r_crec_ingresos_1-3_4-6	0.125
%_ingresos_plan	-0.045
%_ingresos_plan_sinexceso	-0.002
%_ingresos_plan_exceso	0.361
categoria_exceso_ingresos	-0.542
%_ingresos_historico	0.221
%_ingresos_plan_sinexceso_historico	0.344
%_ingresos_plan_exceso_historico	-0.299
categoria_exceso_ingresos_historico	0.648

Attribute	Weight
r_crec_ingresos_1	-0.314
r_crec_ingresos_2	-0.212
r_crec_ingresos_3	-0.619
r_crec_ingresos_4	-0.591
r_crec_ingresos_5	-0.446
r_crec_ingresos_prom	0.176
q_recargas_(t-1)	-0.362
q_recargas_(t-2)	0.253
q_recargas_(t-3)	-0.242
q_recargas_(t-4)	-0.070
q_recargas_(t-5)	0.271
q_recargas_(t-6)	0.546
prom_q_recargas_1-3	0.439
prom_q_recargas_4-6	-0.147
dif_q_recargas_1-3_4-6	0.112
r_crec_q_recargas_1-3_4-6	-0.525
%_q_recargas_historico	-0.009
%_q_recargas_plan_sinexceso_historico	0.080
%_q_recargas_plan_exceso_historico	0.249
categoria_exceso_q_recargas_historico	0.002
r_crec_q_recargas_1	-0.179
r_crec_q_recargas_2	-0.395
r_crec_q_recargas_3	0.010
r_crec_q_recargas_4	0.100
r_crec_q_recargas_5	0.035
r_crec_q_recargas_prom	0.220
n_acc_cobranza_(t-1)	3.167

Attribute	Weight
n_acc_cobranza_(t-2)	-0.510
n_acc_cobranza_(t-3)	-0.384
n_acc_cobranza_(t-4)	-0.450
n_acc_cobranza_(t-5)	0.307
n_acc_cobranza_(t-6)	0.203
prom_n_acc_cobranza_1-3	0.036
prom_n_acc_cobranza_4-6	-0.231
dif_n_acc_cobranza_1-3_4-6	-0.108
r_crec_n_acc_cobranza_1-3_4-6	0.165
%_n_acc_cobranza_historico	0.760
%_n_acc_cobranza_plan_sinexceso_historico	-0.660
%_n_acc_cobranza_plan_exceso_historico	-0.451
categoria_exceso_n_acc_cobranza_historico	-0.182
r_crec_n_acc_cobranza_1	-0.794
r_crec_n_acc_cobranza_2	-0.163
r_crec_n_acc_cobranza_3	-0.135
r_crec_n_acc_cobranza_4	0.104
r_crec_n_acc_cobranza_5	-0.117
r_crec_n_acc_cobranza_prom	0.259
suma_atraso	-0.056
atraso_promedio	-0.070
n_bolsas_(t-1)	-0.498
n_bolsas_(t-2)	-0.183
n_bolsas_(t-3)	0.346
n_bolsas_(t-4)	0.415
n_bolsas_(t-5)	0.712
n_bolsas_(t-6)	0.380

Attribute	Weight
prom_n_bolsas_1-3	-0.520
prom_n_bolsas_4-6	-0.364
dif_n_bolsas_1-3_4-6	0.128
r_crec_n_bolsas_1-3_4-6	0.224
%_n_bolsas_historico	-0.119
%_n_bolsas_plan_sinexceso_historico	-0.270
%_n_bolsas_plan_exceso_historico	0.309
categoria_exceso_n_bolsas_historico	-0.487
r_crec_n_bolsas_1	0.298
r_crec_n_bolsas_2	0.155
r_crec_n_bolsas_3	0.251
r_crec_n_bolsas_4	0.135
r_crec_n_bolsas_5	0.037
r_crec_n_bolsas_prom	-0.138
kb_gprs_(t-1)	1.113
kb_gprs_(t-2)	1.008
kb_gprs_(t-3)	0.002
kb_gprs_(t-4)	-0.497
kb_gprs_(t-5)	0.334
kb_gprs_(t-6)	0.053
prom_kb_gprs_1-3	1.774
prom_kb_gprs_4-6	-1.967
dif_kb_gprs_1-3_4-6	-2.601
r_crec_kb_gprs_1-3_4-6	-1.337
consumo_kb_gprs	0.441