

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA INDUSTRIAL Y DE
SISTEMAS



**“MODELO PARA LA IDENTIFICACIÓN DE FRAUDE BYPASS
INTERNACIONAL EN UNA EMPRESA DE TELEFONIA MÓVIL”**

INFORME DE SUFICIENCIA

PARA OPTAR POR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

PRESENTADO POR:

RONAL NOEL ALVA TARAZONA

LIMA, PERÚ

2013

DEDICATORIA

A mi madre, hermana Paty y mi Esposa Lesly,
Por el apoyo constante y el amor que me dan.

ÍNDICE

RESUMEN EJECUTIVO.....	4
DESCRIPTORES TEMÁTICOS	5
INTRODUCCIÓN.....	6
CAPÍTULO I: PENSAMIENTO SISTÉMICO	7
1. DIAGNÓSTICO FUNCIONAL.....	7
1.1. <i>Productos y/o servicios Movistar Móvil</i>	8
1.2. <i>Clientes</i>	11
1.3. <i>Proveedores</i>	11
1.4. <i>Organización</i>	13
2. DIAGNÓSTICO ESTRATÉGICO	14
2.1. <i>Visión</i>	14
2.2. <i>Historia</i>	14
2.3. <i>Objetivos estratégicos</i>	14
2.4. <i>Análisis FODA</i>	15
2.4.1. <i>Fortalezas y Debilidades para la Lucha contra el Fraude en las Llamadas LDI</i>	15
2.4.2. <i>Oportunidades y Amenazas</i>	15
2.4.3. <i>Análisis de las 5 fuerzas de Porter para el servicio de llamadas LDI</i>	16
CAPÍTULO II: MARCO TEÓRICO.....	17
1. DATA MINING	17
1.1. DEFINICIÓN	17
1.2. METODOLOGÍA.....	17
1.3. PROCESO DE DATA MINING.....	18
1.3.1. CRISP	19
1.3.2. SEMMA.....	21
1.4. TÉCNICAS DE DATA MINING.....	22
1.4.1. ALGORITMOS PREDICTIVOS.....	22
1.4.1.1. ARBOLES DE DECISIÓN	22
1.4.1.2. REDES NEURONALES	23
1.4.1.3. REGRESIÓN LOGÍSTICA.....	25
1.4.1.4. REDES BAYESIANAS	25
1.5. HERRAMIENTA DE ANÁLISIS: SAS MINER	26
CAPÍTULO III: PROCESO DE TOMA DE DECISIONES	28
1. IDENTIFICACIÓN DEL PROBLEMA	28
1.1. <i>Problema Principal:</i>	31
2. PLANTEAMIENTO DE ALTERNATIVAS DE SOLUCIÓN	31
2.1. <i>Alternativa 1: Uso de un software In-House para detectar los clientes fraudulentos.</i>	31
2.2. <i>Alternativa 2: Uso de un modelo de minería de datos para el perfilamiento de clientes fraudulentos.</i>	36
3. SELECCIÓN DE UNA ALTERNATIVA DE SOLUCION	37
3.1. <i>Valoración de alternativas</i>	37
3.2. <i>Evaluación Financiera</i>	38
3.3. <i>Evaluación Cualitativa</i>	39
3.4. <i>Alternativa elegida</i>	40

4.	PLANES DE ACCION PARA DESARROLLAR LA SOLUCION PLANTEADA	40
4.1.	<i>Matriz de datos:</i>	41
4.2.	<i>Selección de la Muestra:</i>	41
4.3.	<i>Exploración:</i>	41
4.4.	<i>Selección de variables:</i>	42
4.5.	<i>Modelamiento:</i>	46
4.6.	<i>Testeo:</i>	46
	CAPÍTULO IV: RESULTADOS	47
1.	RESULTADOS DE LA SOLUCIÓN PLANTEADA.....	47
	CONCLUSIONES Y RECOMENDACIONES	48
1.	CONCLUSIONES	48
2.	RECOMENDACIONES	48
	BIBLIOGRAFÍA.....	49

RESUMEN EJECUTIVO

El presente informe se enfoca en el desarrollo de una solución para detectar uno de los tipos de fraude que más pérdidas ocasiona a las empresas de telefonía móvil, este fraude conocido como "Bypass Internacional" se puede enfrentar desde 3 perspectivas. Desde la perspectiva de la Operadora: a través de la búsqueda de herramientas para la detección del fraude que conlleven a tomar acciones oportunas, desde la perspectiva del Gobierno: a través de las normas y leyes que permitan la sanción de los clientes fraudulentos y desde la perspectiva de la Población: a través de la educación y buen uso del servicio brindado por las empresas de telecomunicaciones. De estas 3 perspectivas la que se desarrolla en este informe es la que le compete a la Operadora. Para ello se plantearon 2 alternativas de solución: La primera se basa en el uso de un software In-House "IL-Hunter" que mediante un robot de llamadas permite identificar clientes que cometen este tipo de fraude. La segunda alternativa se basa en el desarrollo de un modelo de minería de datos para encontrar patrones de comportamiento que nos permitan identificar a los clientes con perfil fraudulento. Finalmente de las 2 alternativas nos quedamos con la segunda por ser una solución más rápida y menos costosa.

DESCRIPTORES TEMÁTICOS

- BY PASS LLAMADAS LDI
- FRAUDE TELEFONIA MOVIL
- INTELIGENCIA DE NEGOCIOS
- MINERIA DE DATOS
- ILEGAL LANDING
- VOZ IP
- TELECOMUNICACIONES

INTRODUCCIÓN

En los últimos 20 años cada vez más las personas a nivel mundial tienen la necesidad de comunicarse con otras personas alrededor del mundo y una de las formas de comunicación más común es a través del uso de la telefonía móvil mediante las llamadas de larga distancia internacional (LDI). Es en este contexto que han aparecido mafias dedicadas al fraude que haciendo un uso indebido del servicio de telefonía móvil se aprovechan de las debilidades que tienen las operadoras para poder detectarlos a tiempo y de los vacíos legales que no permiten sancionarlos debidamente. Dentro de los diferentes tipos de fraude en telefonía móvil vamos a tratar el fraude conocido como "Bypass Internacional" que consiste en el aterrizaje de llamadas vía voz IP desde un país extranjero hasta nuestro país para luego ser redirigido a su destino final utilizando la red local y originando de esta manera pérdidas cuantiosas a los operadores de telefonía móvil por ingresos no percibidos y al estado por evasión de impuestos.

Existen muchos factores que facilitan el cometimiento de fraudes en telecomunicaciones, entre ellos, la evolución tecnológica que se ha logrado en la telefonía móvil y el internet con sus innovadoras aplicaciones, el uso de estos medios ha provocado que se incrementen las posibilidades, la frecuencia, la magnitud de los delitos de fraude que a diario se cometen. Además se debe considerar que los costos de la telefonía internacional son bastante elevados en comparación con la telefonía local.

CAPÍTULO I: PENSAMIENTO SISTÉMICO

1. DIAGNÓSTICO FUNCIONAL

Telefónica es uno de los operadores integrados de telecomunicaciones líder a nivel mundial en la provisión de soluciones de comunicación, información y entretenimiento, con presencia en Europa y Latinoamérica. Tiene operaciones en **25 países** y un promedio de 285,000 empleados. Importe neto de la cifra de negocios (ingresos) de 46.519 millones de euros y más de 313.8 millones de clientes en septiembre de 2012.

En el Perú está presente bajo la marca comercial "**Movistar**" y tiene 4 unidades de Negocio principales:

- 1) Movistar Móvil
- 2) Movistar Fijo
- 3) Internet
- 4) TV Paga

De las 4 unidades de Negocio nos vamos a enfocar en el diagnóstico funcional de la Telefonía Móvil.



Fuente: Intranet de Telefónica

1.1. Productos y/o servicios Movistar Móvil

Los Productos y Servicios más importantes que tiene el Negocio Móvil de Movistar son:

1.1.1. Productos

- Planes Postpago:** Acceso a comunicación vía telefonía celular sin restricciones, a cualquier hora del día, todos los días del año y sin importar donde se encuentre la persona con la que se desea hablar. La mayoría de los planes postpago vienen con bonos de SMS, MMS, Acceso a Internet, Servicio Roaming Internacional activo y otros. El consumo es facturado a un recibo mensual y el servicio brindado está regulado por un contrato de vigencia de 6 meses a más.

- **Planes Prepago:** Ofrece una comunicación libre, rápida, económica, sin contratos y al alcance de todos. Sólo se necesita adquirir un chip y realizar recargas según las necesidades del usuario. Las opciones de recarga son mediante tarjeta, virtual o directamente en establecimientos afiliados.
- **Internet Móvil:** Ofrece navegación desde laptop o PC sin necesidad de cables y a mayor velocidad. Permite acceder a una conexión de banda 2G (700 kbps) y 3G (1500 kbps), con total movilidad bajo toda la cobertura que ofrece la red de Movistar, ya que la navegación se efectúa a través de un módem conectado al puerto USB del equipo de cómputo.
- **Paquetes Voz, SMS e Internet:** Acceso a los servicios de Voz, SMS e Internet de manera adicional a la configuración del plan del cliente.

Grafico: Distribución planta activa x producto – Oct2012

PRODUCTO	NRO ABONADOS
BAM CONTRATO	25,682
BAM CONTROL	155,166
BAM PREPAGO	1,838
CELULAR CONTRATO	2,038,851
CELULAR CONTROL	1,264,052
CELULAR PREPAGO	11,362,427
MOVITALK CONTRATO	1,915
TOTAL	14,849,931

1.1.2. Servicios

- **RPM:** Comunicación con otros Movistar en la Red Privada Movistar usando un numero corto y reduciendo significativamente los el costo de las llamadas que pueden ser ilimitadas dependiendo del plan adquirido.
- **Movitalk:** Comunicación directa a través de la tecnología Push To Talk similar a la Radio pero usando la Red Móvil GSM.
- **Blackberry:** Servicio de datos administrado por RIM que está orientado a la facilidad de acceso a la Oficina Móvil(Correos y Documentos)
- **Roaming Internacional:** Servicio de Larga Distancia Internacional mediante el cual el abonado puede recibir y hacer llamadas usando su propio numero cuando sale del territorio nacional.
- **SMS por cobrar:** El cliente que paga es el que recibe el SMS.
- **SMS Premium:** Servicio de acceso a contenidos de terceros por SMS o Voz, permitiendo una suscripción. Ejemplos: Horoscopos, Chistes, etc.
- **M2M:** Machine To Machine, servicio de transmisión exclusiva de datos para comunicar 2 maquinas.
- **SMID:** Soluciones Móviles Integrales de Datos, servicio de acceso a datos a través de soluciones hechas a medida principalmente para los clientes Empresas.

1.2. Clientes

Movistar clasifica su cartera de clientes en 3 grandes segmentos:

- **Residencial:** Personas Naturales con DNI u otro documento que tienen uno o más productos de Movistar. Representan el segmento con mayor cantidad de abonados (90 %).
- **Negocios:** Personas Jurídicas con RUC que tienen uno o más productos de Movistar. Por lo general son PYMES o empresas Medianas.
- **Empresas:** Son Grandes Empresas que tienen productos de Movistar, representan el segmento con menos cantidad de abonados (2.5 %) pero su facturación por cliente es alta.

Grafico: Distribución planta activa x segmento – Oct2012

SEGMENTO	NRO ABONADOS
Empresas	373,736
Negocios	1,017,479
Residencial	13,458,716
TOTAL	14,849,931

1.3. Proveedores

La posibilidad de inscripción en el Catálogo de Proveedores de Telefónica está disponible para todas aquellas empresas que tengan interés en trabajar con Telefónica. Esa inscripción puede realizarse a través del Portal del Proveedor.

<http://www.telefonica.com/es/suppliers/html/home/index.shtml>.

- **Proveedores de los Servicios de Larga Distancia**

Los proveedores de Movistar para los servicios de Larga Distancia tienen experiencia en el mercado nacional e internacional, ya sea en equipos, en plataformas, en soluciones de Telecomunicaciones, Centralitas, Mantenimiento y reparaciones de Equipos, etc y son los más importantes los siguientes:

- **HUAWEI:**

Empresa China que está en el mercado peruano no más de 2 años, con muy buenos productos y soluciones integradas.

- **CISCO:**

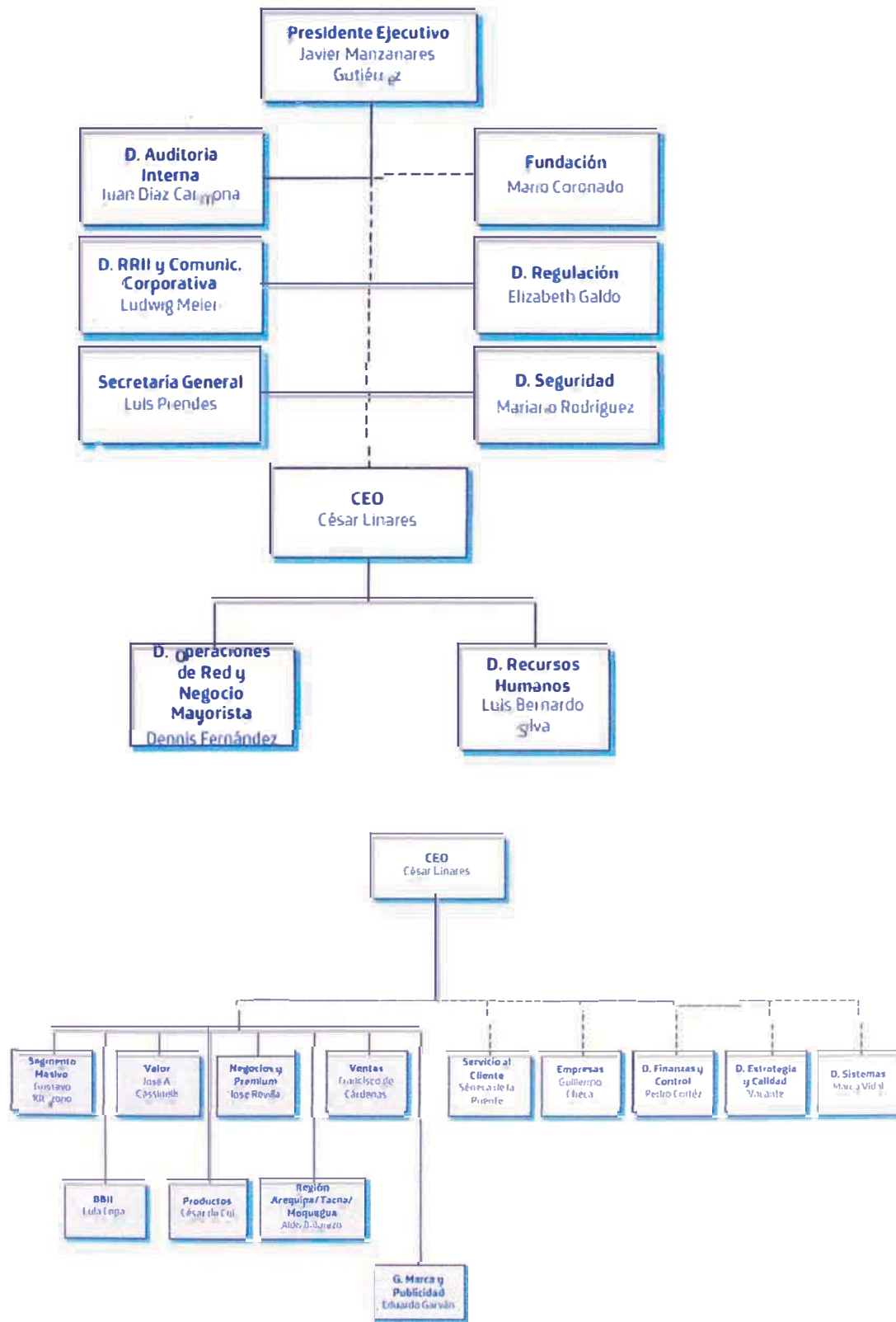
Empresa USA que esta muchos años en el mercado peruano, proporciona equipos, centralitas.

- **ALCATEL:**

Empresa Francesa que se encuentra muchos años en el Perú, es uno de los principales proveedores, principal proveedor de mantenimiento, reparaciones de centrales, también de equipos satelital.



1.4. Organización



Fuente: Intranet de Telefónica

2. DIAGNÓSTICO ESTRATÉGICO

2.1. Visión

Abrimos camino para seguir transformando posibilidades en realidad, con el fin de crear valor para clientes, empleados, sociedad, accionistas y socios a nivel global.

2.2. Historia

1924 - 1954	1965 - 1989	1990 - 1994	1995 - 1999
<ul style="list-style-type: none"> Se constituye en Madrid la CTNE, participada por ITT (1924) El Estado español toma el 79,6% de las acciones de la CTNE (1945) Instala el teléfono 1 millón (1953) La CTNE es ya la 1ª empresa de España (1960) 	<ul style="list-style-type: none"> Inicio de las comunicaciones por satélite (1967) Puesta en servicio de la 1ª Red Especial de Transmisión de Datos de Europa (1971) Instala el teléfono 10 millones (1978) Empieza a cotizar en la Bolsa de Nueva York (1987) 	<ul style="list-style-type: none"> Entra en Chile (CTC) y en Argentina (TASA) (1990) Lanza MovilLine en España (1990), telefonía móvil analógica que logra cubrir el 98% del territorio en tres años. Entra en Perú (TdP) (1994) Lanza la telefonía móvil digital: Movistar (1994) 	<ul style="list-style-type: none"> Inicios del Internet comercial. Telefónica lanza Infovia (1995) Gana el concurso para adquirir Telesp en Brasil (1998) Telefónica se privatiza por completo de nuevo (1999) Lanza el servicio de acceso fijo de banda ancha ADSL (1999)
2000 - 2004	2005 - 2006	2007 - 2008	2009 - 2012
<ul style="list-style-type: none"> Operación Verónica: Telefónica de Argentina; Telefónica del Perú; Telesp y Tele Sudeste en Brasil (2000) Telefónica y Portugal Telecom crean Vivo en Brasil (2003) Lanza Imagenio (2003) Adquisición de los activos de BellSouth en América Latina (2004) 	<ul style="list-style-type: none"> Adquisición de Cesky Telecom (2005) Compra el 5% de China Netcom (2005) Adquisición de los activos de O2 en UK, Alemania e Irlanda (2006) Adjudicación de la licencia de móvil en Eslovaquia (2006) Compra el 51% de Colombia Telecom (2006) 	<ul style="list-style-type: none"> Alianza industrial con Telecom Italia (10,47% de los derechos de voto) (2007) Adquisición de Telemig por Vivo en Brasil (2008) Alcanza una participación del 5,38% en China Unicom tras la fusión CU-CNC (2008) Incremento de la participación en Telefónica CTC Chile hasta el 97,89% (2008) 	<ul style="list-style-type: none"> Refuerzo de la alianza estratégica con China Unicom (2009) Adquisición de Hansenet en Alemania, Jajah en Israel y Tuenti en España (2010) Toma de control de Vivo mediante la compra a PT de su participación en Brasilcel (2010) Adjudicación de la licencia de móvil en Costa Rica (2011) Venta del 4,56% de China Unicom (2012)

Fuente: <http://www.telefonica.com>

2.3. Objetivos estratégicos

- La maximización del valor de la Compañía en interés de los accionistas:
- Contribuir al desarrollo social, económico y tecnológico de las comunidades.
- Incremento de los ingresos provenientes del acceso a Internet Móvil.

2.4. Análisis FODA

El proceso de planeación estratégica comienza por un análisis interno y el entorno externo (5 Fuerzas de Porter para un entorno competitivo) el cual es importante realizar y para ello vamos a utilizar el análisis FODA (Análisis de las Fortalezas, Debilidades, oportunidades y amenazas) en el **servicio de Llamadas de Larga Distancia Internacional**.

2.4.1. Fortalezas y Debilidades para la Lucha contra el Fraude en las Llamadas LDI

FORTALEZAS:	DEBILIDADES:
<ul style="list-style-type: none"> • Especialistas en Lucha contra el Fraude Móvil, con más de 10 años de experiencia en el tema. • Especialistas en Minería de datos. • Compromiso de la Gerencia de Riesgo Operacional y la Gerencia de Inteligencia de Negocios para combatir el Fraude. 	<ul style="list-style-type: none"> • Vacíos legales que no permiten sancionar drásticamente a las personas o empresas que realizan los actos de fraude. • No se cuenta con un sistema especializado en detección de Fraude en Telefonía Móvil. • Demora en la detección de los actos de Fraude.

2.4.2. Oportunidades y Amenazas

OPORTUNIDADES:	AMENAZAS:
<ul style="list-style-type: none"> • Adquisición de un sistema especializado en detección de Fraude en Telefonía Móvil. • Desarrollo de un Modelo de Minería de datos para detectar patrones de comportamiento de los clientes fraudulentos. 	<ul style="list-style-type: none"> • Perdidas hasta por 2.6 millones de dólares producto de los ingresos no percibidos • Humanización del comportamiento del tráfico de llamadas para evitar ser detectado.

2.4.3. Análisis de las 5 fuerzas de Porter para el servicio de llamadas LDI.



CAPÍTULO II: MARCO TEÓRICO

1. DATA MINING

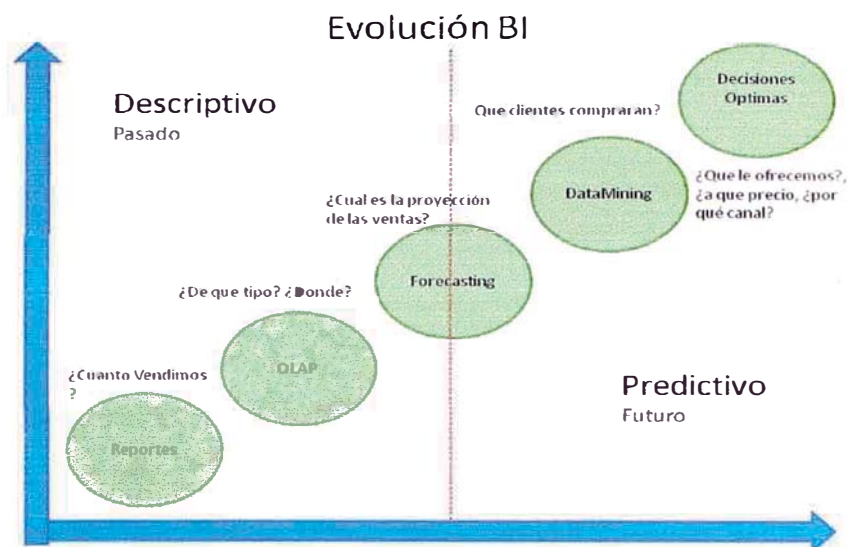
1.1. Definición

Gartner Group: “Data Mining es el proceso de descubrir nuevas y significantes correlaciones, patrones y tendencias en grandes cantidades de datos almacenados en repositorios usando tecnologías de reconocimiento de patrones así como técnicas estadísticas y matemáticas”.¹

1.2. Metodología

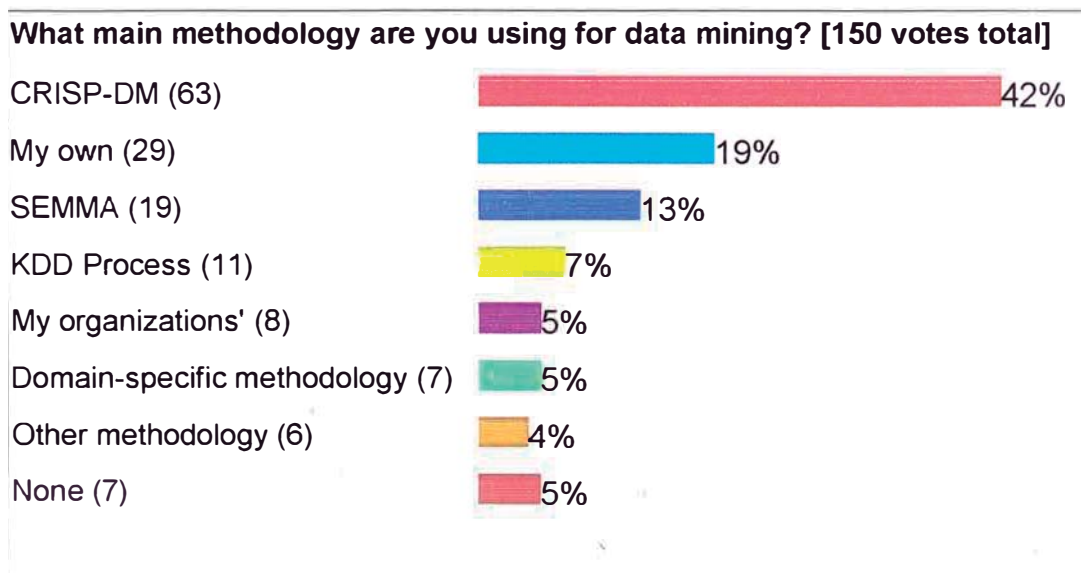
Las técnicas de Data Mining son el resultado de un largo proceso de investigación y desarrollo de productos (I+D). Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. Data Mining toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva.

¹ Gartner Group. *Data Mining* [en línea]: *SAMBA department*.
<[http://www2.nr.no/documents/samba/research_areas/BAMG/Pattern/Data Mining.html](http://www2.nr.no/documents/samba/research_areas/BAMG/Pattern/Data%20Mining.html)>.
[Consulta 18 Noviembre 2012].

Gráfico: Evolución de BI²

1.3. Proceso de Data Mining

Los procesos más aceptados según Kdnuggets son CRISP y SEMMA.

Gráfico: Cual es la principal metodología que Ud. Usa para Data Mining³

² Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas]. Lima Perú: Escuela de Postgrado de la UPC, 2012.

³ Kdnuggets. *What main methodology are you using for data mining?* [en línea]: <http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm>. [Consulta 18 Noviembre 2012].

1.3.1. CRISP⁴

Cross – Industry Standard Process for Data Mining, financiado por la Comunidad Europea y el esfuerzo de consorcios de empresas europeas: NCR(Dinamarca), AG(Alemania), SPSS(Inglaterra) y OHRA(Holanda).

Características:

- No propietario.
- Independiente de la aplicación o la industria.
- Neutral con respecto a herramientas.
- Enfocado en problemas del negocio así como en el análisis técnico.

Fases:

1. Comprensión del negocio o problema.
2. Comprensión de los datos.
 - ✓ Recolección de los datos
 - ✓ Descripción de los datos
 - ✓ Exploración de los datos
 - ✓ Verificación de calidad
3. Preparación de los datos.
 - ✓ Selección de los datos
 - ✓ Limpieza de los datos
 - ✓ Estructuración de los datos
 - ✓ Integración de los datos
 - ✓ Formato de los datos

⁴ Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas]. Lima Perú: Escuela de Postgrado de la UPC, 2012.

4. Modelado

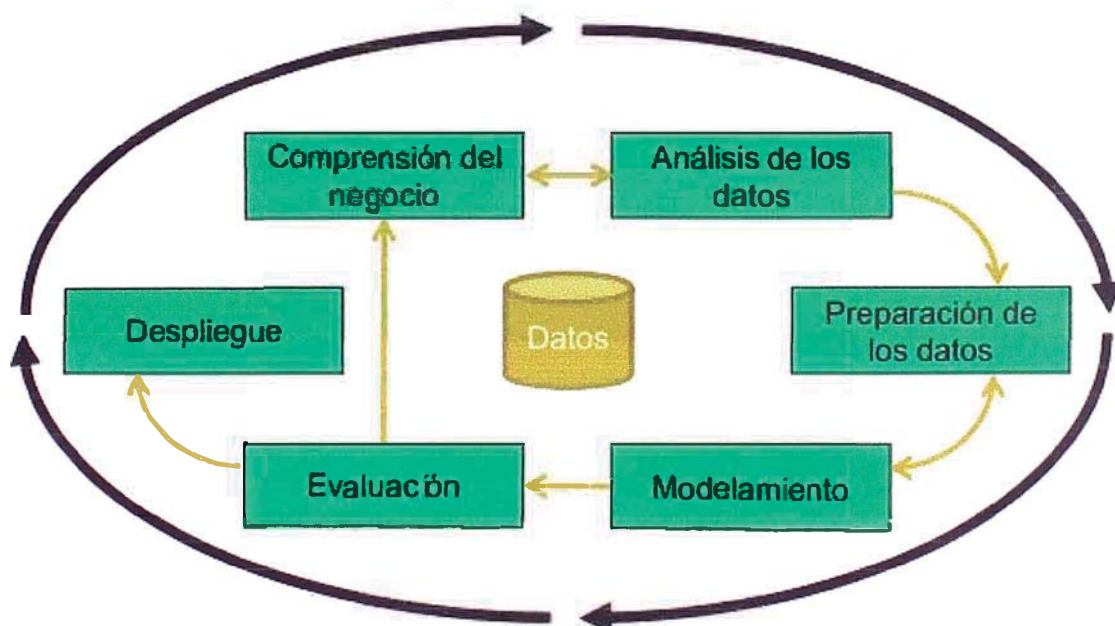
- ✓ Selección de la técnica de modelado
- ✓ Generación del plan de prueba
- ✓ Construcción del modelo
- ✓ Evaluación del modelo

5. Evaluación

- ✓ Evaluación de los resultados
- ✓ Proceso de revisión
- ✓ Determinación de futuras fases

6. Implementación

- ✓ Plan de implementación
- ✓ Informe final

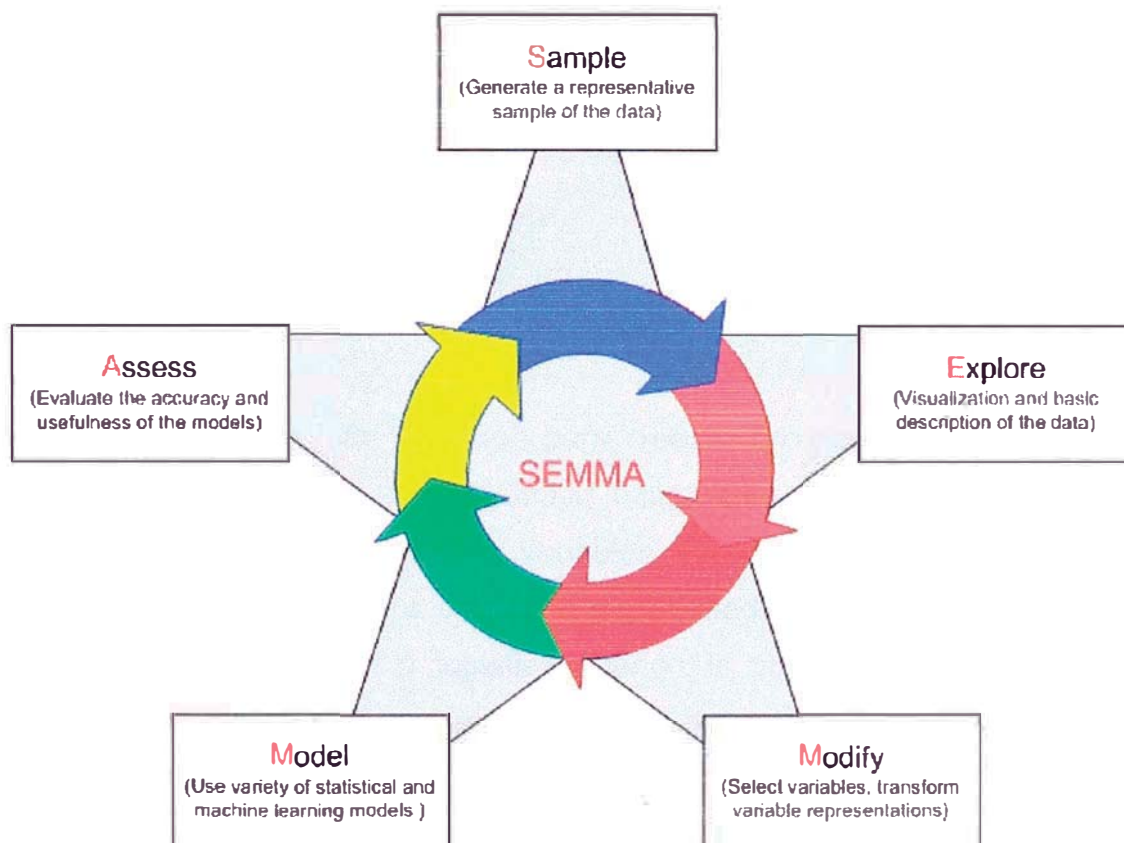


1.3.2. SEMMA⁵

“Sample, Explore, Modify, Model, Assess”, es un proceso presentado por SAS Institute e implementado en su herramienta SAS Enterprise Miner.

Fases:

1. Muestreo (Sample)
2. Exploración (Explore)
3. Manipulación (Modify)
4. Modelado (Model)
5. Valoración (Assess)



⁵ Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas]. Lima Perú: Escuela de Postgrado de la UPC, 2012.

1.4. Técnicas de Data Mining

1.4.1. Algoritmos Predictivos

1.4.1.1. Árboles de decisión

“Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema”⁶.

Existen varias técnicas para construir los arboles:

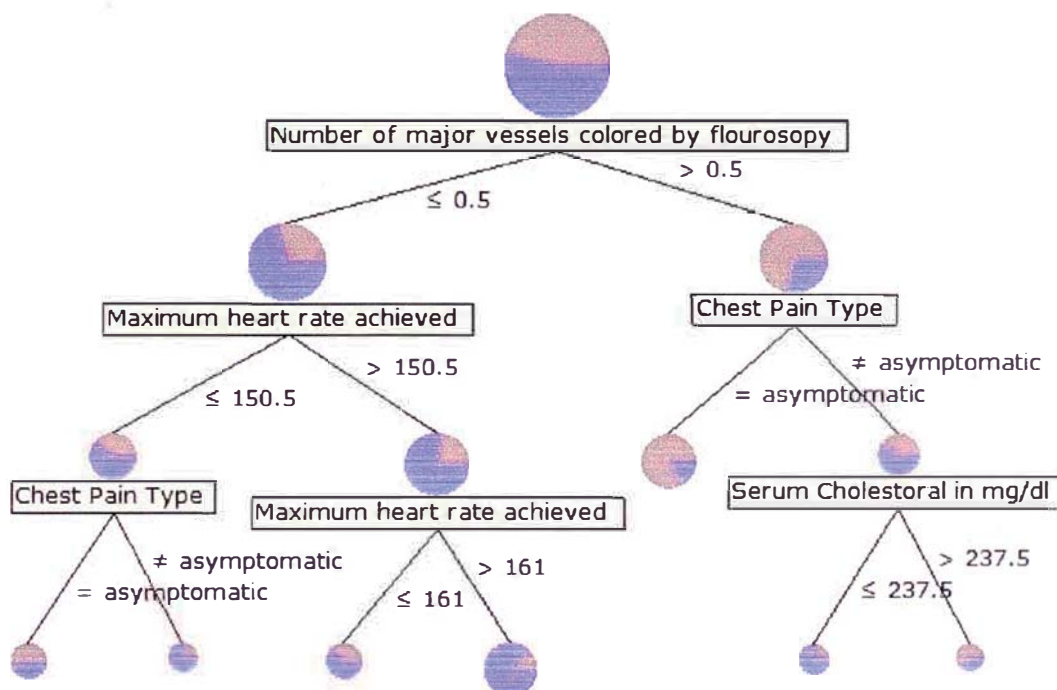
- Regresión Lineal
- Chaid
- C&RT
- Quest

El método más conocido para crearse arboles de decisión es el Chaid (Chi-square Automatic Interaction Detection). Examina la relación entre muchas variables categóricas o discretas y un objetivo categórico o medida de resultado. Es una alternativa a la regresión lineal o logística cuando el conjunto de datos no es adecuado para realizar un análisis de regresión. Realiza pruebas para verificar cual variable explicativa es más significativa para la respuesta.

⁶ WIKIPEDIA. *Minería de datos* [en línea]:
<http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos>. [Consulta 18 Noviembre 2012].

Ventajas del uso de los arboles de decisión⁷:

- De fácil entendimiento
- El resultado es presentado en forma de un diagrama que muestra las variables explicativas que resultan en una mayor diferencia de la variable objetivo
- Trata valores nulos.



1.4.1.2. Redes Neuronales

“Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida”⁸.

⁷ Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas]. Lima Perú: Escuela de Postgrado de la UPC, 2012.

⁸ WIKIPEDIA. *Minería de datos* [en línea]:

<http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos>. [Consulta 18 Noviembre 2012].

La Redes Neuronales son utilizadas cuando hay bastante información pero no es fácil derivar reglas mediante otros algoritmos.⁹

Tipos de Redes Neuronales:

- **MLP – Perceptron Multi Capa:** Consiste en capas de neuronas, donde cada neurona está unida a otras neuronas en la capa anterior por conexiones de ponderación variable. Consisten en una capa de entrada, una capa de salida y, al menos, una capa oculta.

Ventajas:

- Es efectivo en una amplia variedad de problemas.
- Es capaz de generalizar bien.

Desventajas:

- Entrenarlo puede llevar mucho tiempo.
 - No garantiza la mejor solución global.
-
- **RBF – Función Base Radial:** Método alternativo al MLP para hacer ajuste a funciones no lineales. Esta construido usando un sistema rígido de 3 capas y su entrenamiento es hacia adelante.

⁹ Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas]. Lima Perú: Escuela de Postgrado de la UPC, 2012.

1.4.1.3. Regresión Logística

“En estadística, la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados (GLM por sus siglas en inglés) que usa como función de enlace la función logit. Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística. La regresión logística es usada extensamente en las ciencias médicas y sociales.”¹⁰

1.4.1.4. Redes Bayesianas

“Una **red bayesiana**, o **red de creencia**, es un modelo probabilístico multivariado que relaciona un conjunto de variables aleatorias mediante un grafo dirigido que indica explícitamente influencia causal. Gracias a su motor de actualización de probabilidades, el Teorema de Bayes, las redes bayesianas son una herramienta extremadamente útil en la estimación de probabilidades ante nuevas evidencias”¹¹.

¹⁰ WIKIPEDIA. *Minería de datos* [en línea]: <http://es.wikipedia.org/wiki/Regresi%C3%B3n_log%C3%ADstica>. [Consulta 18 Noviembre 2012].

¹¹ P. Naim, P. Wuillemin, P. Leray, O. Pourret, A. Becker. Les réseaux bayésiens, Eyrolles 2004.

1.5. Herramienta de Análisis: SAS MINER

SAS Enterprise Miner es una herramienta que agiliza el proceso de minería de datos para crear modelos predictivos y descriptivos de alta precisión basado en grandes volúmenes de datos de toda la empresa. Ofrece un conjunto de capacidades integradas para crear y compartir puntos de vista que se pueden usar para tomar mejores decisiones.¹²

Beneficios

- Apoyar el proceso de minería de datos de manera completa con un amplio conjunto de capacidades.
- Construir mejores modelos con una mesa de trabajo de minería de datos versátil.
- Permitir a los analistas de negocio de forma rápida y fácilmente obtener conocimientos de una manera autosuficiente y automatizada.
- Mejorar la precisión de las predicciones y compartir fácilmente información fiable que aumente la calidad de las decisiones.
- Facilitar el proceso de implementación y modelo de puntuación.

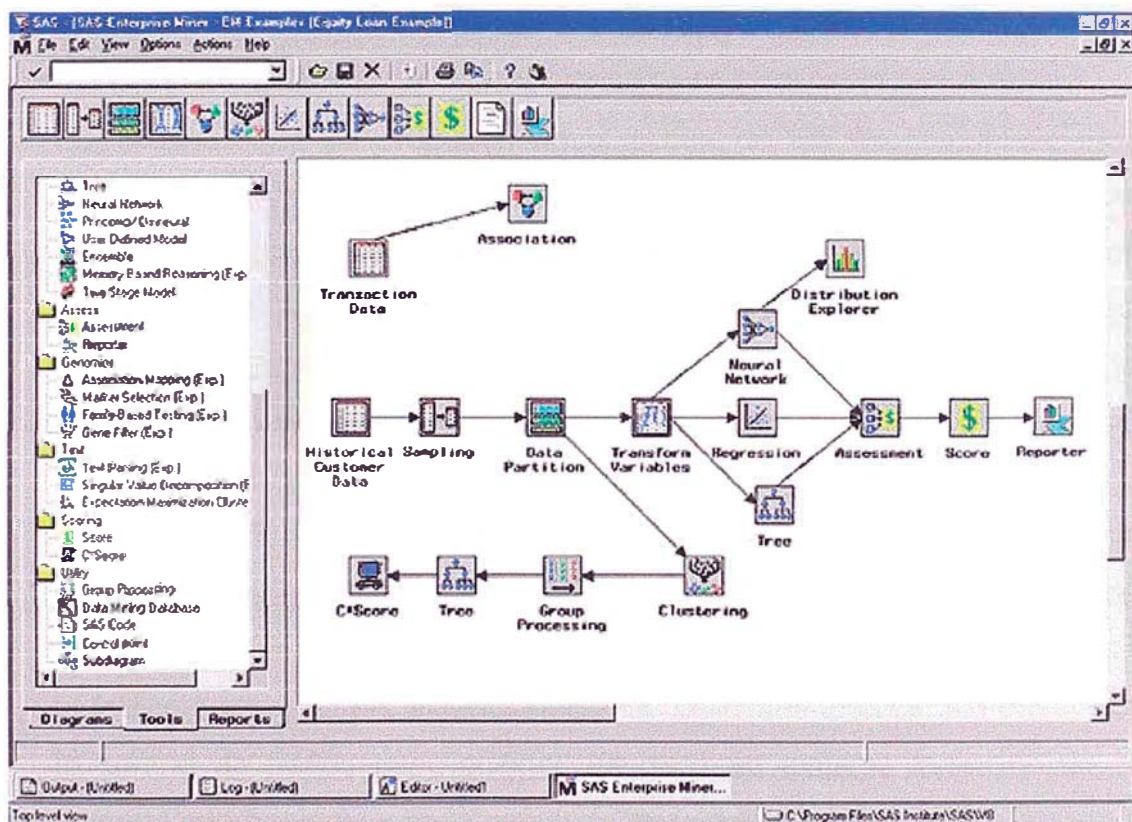
Características

- Potente, fácil de usar interfaz gráfica de usuario, así como el procesamiento por lotes para grandes trabajos
- Preparación de datos, el resumen y la exploración
- Modelado avanzado predictivo y descriptivo
- Manera rápida, fácil y autosuficiente de generar modelos para usuarios de negocios.

¹² SAS. *SAS Enterprise Miner* [en línea]: <
<http://www.sas.com/technologies/analytics/datamining/miner/>>. [Consulta 18 Noviembre 2012].

- Negocios basados en comparaciones de modelos, información y gestión.
- Proceso de calificación automática
- Diseño abierto y extensible
- Procesamiento escalable

Grafico Herramienta SAS Enterprise Miner



CAPÍTULO III: PROCESO DE TOMA DE DECISIONES

1. IDENTIFICACIÓN DEL PROBLEMA

Los fraudes que se producen en el sector de las telecomunicaciones causan cuantiosas pérdidas económicas a las empresas telefónicas en nuestro país y, de similar forma al resto del mundo. Pese a las medidas adoptadas por los Organismos competentes en nuestro país, aun no se logra la aprobación de una legislación penal apropiada que permita desestimular y penalizar adecuadamente estos hechos.

De los tipos de fraude existentes, uno de los que más perjuicio origina a las operadoras de telefonía y al estado Peruano es el "ByPass", que de forma resumida se puede decir que consiste en encaminar directamente el tráfico que viene del exterior hacia las centrales locales, sin pasar por la central de tráfico internacional, es decir se evita la tarificación de la llamada internacional y se la convierte en una llamada local.

El tráfico de las llamadas entrantes es de aproximadamente 8 veces mayor que el de las llamadas salientes y es en este mismo sentido que se cometen el ilícito del ByPass, mostrándose como una ruta alternativa para los carriers internacionales, el cual presenta un costo sumamente menor que el exigido por las compañías telefónicas locales; por este motivo deciden ingresar sus volúmenes de tráfico mediante esta vía alternativa.

El tráfico telefónico es digitalizado y enrutado por internet hasta un local clandestino del país destino, local que previamente debe estar equipado con líneas telefónicas, estas líneas son conseguidas a través de cómplices en la misma empresa telefónica o con documentación falsa.

Una vez que todo el volumen de tráfico ha ingresado al local clandestino, equipos de telecomunicaciones procesan la información como una mini central telefónica; las llamadas procesadas por la mini central telefónica generan llamadas locales hacia los abonados finales, completando así la llamada que se generó desde cualquier parte de mundo hacia nuestro país. Y finalmente las empresas telefónicas locales sólo perciben una llamada local, mientras la porción internacional la cobra la empresa que comete el fraude.

Veamos un ejemplo, para llamadas desde Panamá a Perú.

Comparación Costo Llamada Internacional vs Llamada Local

Tipo Llamada	Origen	Destino	Tarifa x Minuto (Moneda País Origen)	Tipo Cambio (Moneda País Origen / Dólar U.S.A.)	Tarifa x Minuto (Dolares U.S.A.)
LDI	Panamá	Perú	0.46 (1)	1	0.460
Local	Perú	Perú	0.27 (2)	2.6	0.104
Diferencia llamada LDI – Local					0.36

(1) Valor tomado de la página Web del Operador Digicel Panamá, tarifa LDI con destino Perú para líneas Postpago, al 19 Nov 2012.

http://www.digicelpanama.com/es/services/larga_distancia

(2) Valor tomado de la página Web de Movistar Peru, tarifa llamada local de Movistar a Movistar para líneas Postpago "Plan Postpago RPM S/. 319.90 l".

<http://www.movistar.com.pe/movil/postpago/planes-y-tarifas/planes-rpm>

Del cuadro anterior se comprueba que los costos de las llamadas LDI son mucho mayores y en este caso se cuadruplica en relación a los costos de las llamadas locales, es por eso que algunas mafias dedicadas al fraude se aprovechan de esta situación y se dedican al ByPass LDI.

Ingresos percibidos por el Operador Destino según tipo de llamada LDI o Local:

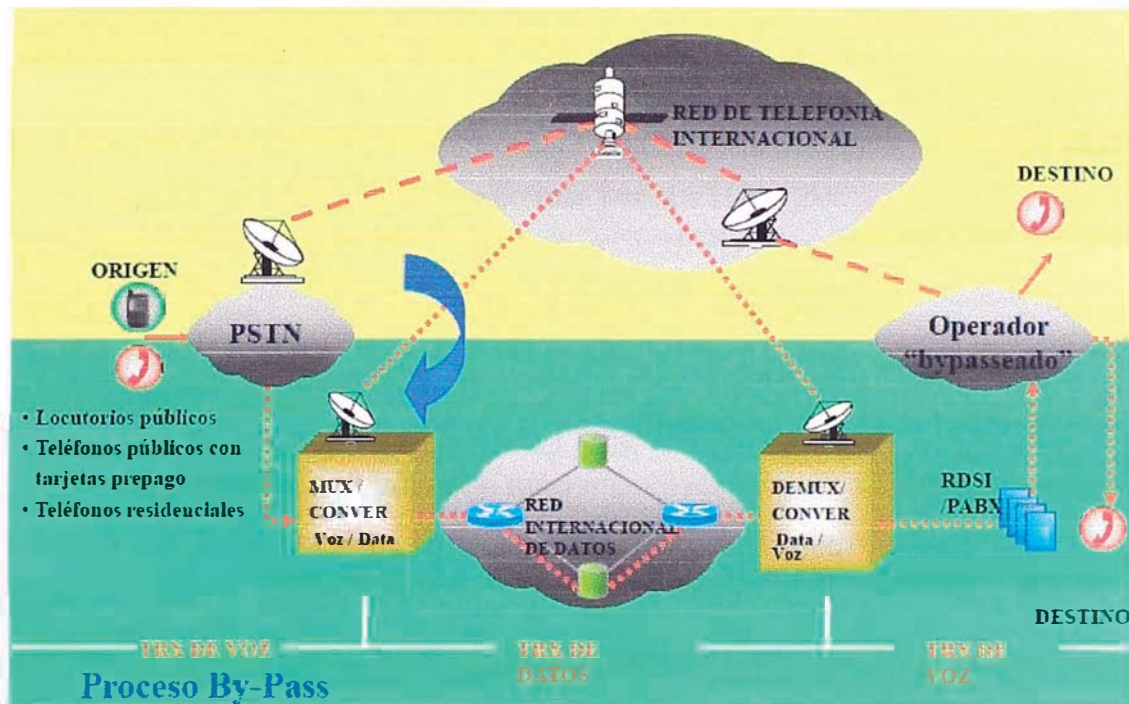
Canal	Origen	Destino	Tipo Llamada	Ingresos por interconexión entrante (Dólares U.S.A)
Autorizado	Panamá	Perú	LDI	(*)0.250
ByPass NO autorizado	Panamá	Perú	Local	0.104
			Ingreso no percibido	0.146

(*) Cargo por interconexión LDI Entrante del Operador Digicel Panamá a Movistar Perú.

Según el cuadro el Ingreso no Percibido por Movistar por cada minuto de una llamada LDI que se realiza a través de ByPass desde Panamá es de 0.146 dólares.

Asumiendo que 0.146 dólares es el valor promedio del ingreso no percibido por cada minuto desde cualquier país origen, al multiplicarlo por 1.5 millones de minutos que es la cantidad promedio de minutos entrantes que provienen de ByPass en un mes, 1.5 millones de minutos, el ingreso no percibido total es de 219,000 dólares por mes. Monto que el estado también deja de recaudar correspondiente al IGV.

Grafico Proceso ByPass Internacional¹³



1.1. Problema Principal:

Detección oportuna de los clientes que realizan fraude ByPass.

2. PLANTEAMIENTO DE ALTERNATIVAS DE SOLUCIÓN

2.1. Alternativa 1: Uso de un software In-House para detectar los clientes fraudulentos.

Nombre del Software: IL-Hunter

Derechos de Autor: Ronal Alva

Tipo de Software: Aplicación de escritorio.

¹³ SUPERTEL. *Fraude en las telecomunicaciones*. Quito Ecuador: Dirección general de investigación especial en telecomunicaciones, 2009.

Descripción del Sistema IL-Hunter: Es una Aplicación de escritorio que permite identificar el fraude ByPass (Local o LDI) a través de un robot de llamadas que puede programar la ejecución de cientos de llamadas usando el saldo de tarjetas prepagadas de países extranjeros. Para lo cual utiliza 2 terminales celulares que se conectan a la PC y sirven para realizar las pruebas completando la traza desde el origen hasta el destino y ubicando el celular intermedio que realiza el fraude.

Dentro de las funcionalidades del sistema tenemos:

- Test de conexión de los terminales origen y destino. Los celulares deben tener función de módem.
- Administración y configuración de los parámetros de las tarjetas prepagadas: Se configuran los datos de una tarjeta como el numero de acceso, PIN, y los tiempos de conexión.
- Administración de pruebas de llamadas: Permite programar la ejecución de cientos de llamadas utilizando las tarjetas configuradas.
- Reportes.

Pantalla de Inicio:



Test de Conexión:

IL Hunter: Conexión de terminales

Archivo Administración Reportes

IL-Hunter | **Illegal Landing**

Usuario: GP\ralvat

Probar Conexión

Datos Origen:		Datos Destino:	
Fabricante:	Motorola CE, Copyright 2000	Fabricante:	Motorola CE, Copyright 2000
Modelo:	L6i	Modelo:	I7
Puerto:	COM4	Puerto:	COM6

Idioma de prueba:

Utilizar terminal origen Teléfono:

Utilizar terminal destino

Configuración de tarjetas:

IL Hunter: Administración de tarjetas

Archivo Administración Reportes

IL-Hunter | **Illegal Landing**

Usuario: GP\ralvat

Id	Nombre	País	Numero acceso	Activo
1	Tower New York ...	EE.UU.	6465029491	<input checked="" type="checkbox"/>
2	Tower New York ...	EE.UU.	7189071422	<input checked="" type="checkbox"/>
3	Tower New York ...	EE.UU.	3472899668	<input checked="" type="checkbox"/>
4	Tower Fairfield	EE.UU.	2038440621	<input checked="" type="checkbox"/>
5	Aficionado New ...	EE.UU.	2014312688	<input checked="" type="checkbox"/>
6	Aficionado New ...	EE.UU.	2012542007	<input checked="" type="checkbox"/>
10	La Mia Hackensack	EE.UU.	2012213240	<input checked="" type="checkbox"/>
11	La Mia Jersey City	EE.UU.	2012344026	<input checked="" type="checkbox"/>

Datos de la Tarjeta:

Nro tarjeta: 10 Teléfono Destino: 51198499399 (Cod País + Cod. Ciudad + Teléfono)

Nombre: La Mia Hackensack

País: EE.UU. Código: 1


Nro acceso: 2012213240

Idioma: 1 Tiempo: 13

Pin: 5821653809# Tiempo: 10

Cod. inter.: 011 Tiempo: 37

Activa: Tiempo conexión llamada: 59

Administración de Pruebas:

IL Hunter. Administración de pruebas

Archivo Administración Reportes

IL-Hunter | **Illegal Landing**

Usuario: GP\valvat

Id	Nombre	Tipo_prueba	Activo
3	Secuencial EE.UU.	Secuencial	<input checked="" type="checkbox"/>
4	Intercalado EE.UU.	Intercalado	<input checked="" type="checkbox"/>
5	La Mia EE.UU. Secuencial	Secuencial	<input checked="" type="checkbox"/>
7	prueba	Secuencial	<input checked="" type="checkbox"/>

Datos de la Prueba:

Nro. Prueba: 6

Nombre:

Activo:

Tipo prueba:

Secuencial

Intercalado

Tarjetas:

Id	Nombre	País	SeL.
1	Tower New York City 1	EE.UU.	<input type="checkbox"/>
2	Tower New York City 2	EE.UU.	<input type="checkbox"/>
3	Tower New York City 3	EE.UU.	<input type="checkbox"/>
4	Tower Fairfield	EE.UU.	<input type="checkbox"/>
5	Accionado New Jersey - Cliffside Park	EE.UU.	<input type="checkbox"/>

Ejecución de pruebas:

IL Hunter. Ejecución de pruebas

Archivo Administración Reportes

IL-Hunter | **Illegal Landing**


Usuario: GP\valvat

Llamada de prueba:

Llamar origen

Llamar destino

Telefono:





Selecciona una prueba:

Id	Nombre	Tipo_prueba
3	Secuencial EE.UU.	Secuencial
4	Intercalado EE.UU.	Intercalado
5	La Mia EE.UU. Secuencial	Secuencial
7	prueba	Secuencial

Iteraciones x tarjeta:

Total iteraciones:

Reportes:

IL Hunter: Reporte de detecciones

Archivo Administración Reportes

IL Hunter Illegal Landing

Usuario: GPVrvalvet

Lineas detectadas

Nro	Ejec	Prueba	Tarjeta	Pais	Estado	Mod	Origen	Numero_acceso	Llam. Origen	Telefono
185	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:22 a.m.	'54959499000'
186	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:24 a.m.	'1999288493'
187	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:26 a.m.	'1975179542'
188	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:28 a.m.	'54959499000'
189	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:30 a.m.	'1988052569'
190	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:32 a.m.	'1975179542'
191	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:34 a.m.	'1988052397'
192	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:36 a.m.	'54959574206'
193	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:38 a.m.	'1988052569'
194	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:40 a.m.	'1988052468'
195	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:42 a.m.	'1988052569'
196	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:45 a.m.	'1996034557'
197	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:47 a.m.	'1988052475'
198	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:49 a.m.	'54959574206'
199	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:51 a.m.	'1988052397'
200	35	La Mía EE.U...	La Mía Jerse...	EE.UU.	Entrante	Voz	990073176	2012344026	04/07/2009 12:53 a.m.	'54959574206'

Ventajas:

- 100% de certeza en la detección de los clientes que realizan fraude ByPass.

Desventajas:

- No se puede barrer a toda la base de clientes.
- El costo es alto para alcanzar un gran número de detecciones, es decir el número de detecciones es directamente proporcional al costo, mientras más tarjetas se compren hay más posibilidad de detectar a los clientes fraudulentos.

2.2. Alternativa 2: Uso de un modelo de minería de datos para el perfilamiento de clientes fraudulentos.

Mediante técnicas de Data Mining se ha desarrollado un modelo predictivo para la detección de clientes con el perfil fraudulento ByPass LDI.

Técnica:	Predicción
Algoritmo:	Arboles de decisión:
Metodología:	SEMMA
Herramienta:	SAS Enterprise Miner Client 6.1.

Ventajas:

- Se puede barrer el total de la base de clientes para encontrar los que tienen perfil fraudulento.
- El costo es mucho menor porque Movistar ya tiene usa el SAS Enterprise Miner en el área de BI y cuenta con licencias.
- La elaboración del modelo no toma mucho tiempo, toma de 2 a 3 semanas y luego la réplica a el total de la base toma solo horas.

Desventajas:

- Nivel de certeza menor al 100%, pero por ser un modelo de fraude el Lift esperado es mayor a 90 %.

3. SELECCION DE UNA ALTERNATIVA DE SOLUCION

3.1. Valoración de alternativas

Es una evaluación de las alternativas basado en juicio de expertos y donde se analizan los pros, contras, viabilidad, tiempos. Está basado en la experiencia.

Alternativa	1	2
Pros	<i>Nivel de certeza 100%</i>	<i>Se puede analizar toda la base de clientes.</i>
Contras	<i>No se puede analizar toda la base de clientes. Nro de detecciones es directamente proporcional al costo de inversión en las tarjetas prepagadas y al número de llamadas programadas.</i>	<i>Nivel de certeza menor al 100%, pero por encima del 90%</i>
Viabilidad	<i>Viable</i>	<i>Viable</i>
Razones	<i>Para detectar casos de fraude comprobados.</i>	<i>Técnicas potentes y modernas que utilizan actualmente las grandes empresas alrededor del mundo con buenos resultados en la aplicación de las áreas de detección de fraude.</i>
Plazo	<i>Corto Plazo</i>	<i>Corto plazo</i>
Escenario optimista	<i>El 100 % de las llamadas programadas culminen el flujo completo y se detecten los clientes fraudulentos.</i>	<i>El nivel de predicción del modelo sea superior al 95 %.</i>
Escenario pesimista	<i>Las llamadas programadas no culminan el flujo completo y no se puede detectar los clientes fraudulentos.</i>	<i>Las variables utilizadas en el modelo no son suficientes para predecir el comportamiento de los clientes fraudulentos.</i>

Según el análisis de valoración de alternativas ambas alternativas son viables.

3.2. Evaluación Financiera

Evaluación desde el punto de vista del ahorro probable que representa cada una de las alternativas.

Alternativa 1: Uso de un software In-House para detectar los clientes fraudulentos.

Tarjetas prepagadas \$5 c/u(100 tarjetas): 500 dólares = **1,295 soles** (*)

Costo envío de tarjetas prepagadas desde el extranjero = **500 soles**

2 terminales celulares = **600 soles**

Saldo para llamadas LDI 2,000 dólares = **5,180 soles** (*)

Personal para configuración de tarjetas y pruebas = 3,500 soles

Costo total x Mes Alter. 1 = 11,075 soles

(*)Tipo Cambio 2.59 soles x dólar, a la fecha del 28/11/2012

Alternativa 2: Uso de un modelo de minería de datos para el perfilamiento de clientes fraudulentos.

Costo Licencia SAS Enterprise Miner: 224,500 soles por año, el costo asociado al modelo sería dividido entre 12 meses y ente la cantidad aprox de modelos que se realizan al mes (aprox 4).

Costo licencia = $224,500 / (12 * 4) = 4,678$ soles

Personal para la elaboración del modelo = **3,500 soles**

Costo total x Mes Alter. 2 = 8,178 soles

Según el análisis de ahorro probable en costos, la alternativa 2 sería más conveniente.

3.3. Evaluación Cualitativa

Criterios y Pesos:

Criterio	Peso %	Descripción
Alcance	30	Aplicación de la alternativa sobre el total de la base de clientes.
Bajo Costo	20	Ahorro probable de c/alternativa.
Nivel de Certeza	30	Nivel de certeza de la detección.
Corto Plazo	10	Que tan fácil es desarrollar la solución en el corto plazo (3 semanas a 1 mes).
Soporte y/o mantenimiento	10	Que tan fácil de mantener en el tiempo es la solución.

Puntuación:

Puntaje	Significado
1	Muy Bajo
2	Bajo
3	Regular
4	Alto
5	Muy Alto

Criterio	Peso %	Calificación	
		Alternativa 1	Alternativa 2
Alcance	30	3	5
Bajo Costo	20	4	5
Nivel de Certeza	30	5	4
Corto Plazo	10	4	4
Soporte y/o mantenimiento	10	5	4

Ponderación alternativa 1: Uso de un software In-House para detectar los clientes fraudulentos = **4.1**

Ponderación alternativa 2: Uso de un modelo de minería de datos para el perfilamiento de clientes fraudulentos = **4.5**

3.4. Alternativa elegida

Debido al mayor puntaje obtenido en la evaluación cualitativa y de costos optamos por la ALTERNATIVA 2.

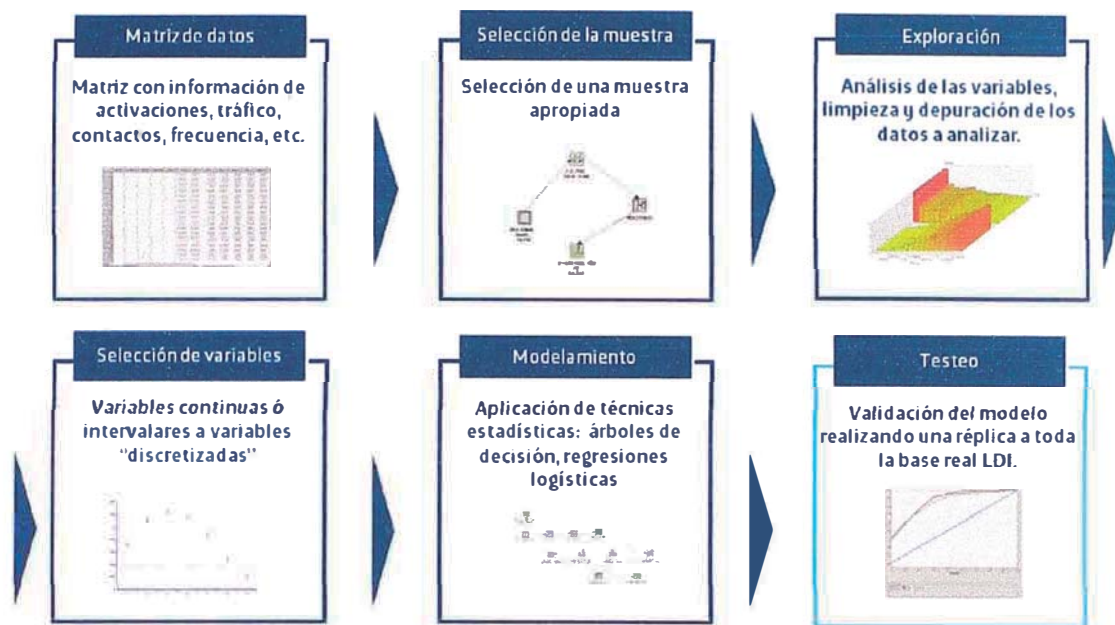
4. PLANES DE ACCION PARA DESARROLLAR LA SOLUCION PLANTEADA

Para el desarrollo del modelo de minería de datos vamos a utilizar la Metodología SEMMA que viene integrada con el software SAS Miner y usaremos el algoritmo predictivo de arboles de decisión.

Características

- Población Objetivo: Líneas celulares Post-pago quienes presentaron tráfico valorizado en el mes de Octubre'10.
- Periodo de análisis: 03 meses (Ago12 - Oct12).
- Universo de variables: información del tráfico de llamadas salientes y entrantes

Etapas del proceso:



4.1. Matriz de datos:

Se generó una matriz de datos con 150 variables de entrada para el desarrollo del modelo.

4.2. Selección de la Muestra:

La muestra aleatoria y representativa fue de 77.5 K abonados y se realizó un balanceo en el que se incluyó 1.8 k abonados identificados como fraudulentos para mejorar los umbrales. Además el 70% de la muestra se utilizó para el entrenamiento y el 30% para validación.

4.3. Exploración:

Se realizó un análisis, limpieza, validación y depuración de las variables donde solo se consideraron variables de tráfico valorizado (voz y sms) saliente y entrante que tienen relación con el target.

4.4. Selección de variables:

Se consideraron 90 variables de tráfico:

Cod	Variables	Tipo Tráfico	Unidad	Tipo Llamada
V1	Nro total minutos RPM	Saliente	Minuto	RPM
V2	Nro total minutos Móvil-Móvil	Saliente	Minuto	OnNet
V3	Nro total llamadas Móvil-Móvil	Saliente	Llamada	OnNet
V4	Nro total destinos distintos Móvil-Móvil	Saliente	Cantidad	OnNet
V5	Promedio total minutos Móvil-Móvil	Saliente	Llamada	OffNet
V6	Mínima duración minutos Móvil-Móvil	Saliente	Minuto	OnNet
V7	Máxima duración minutos Móvil-Móvil	Saliente	Minuto	OnNet
V8	Flag tráfico Móvil-Móvil Mes1	Saliente	Flag	OnNet
V9	Flag tráfico Móvil-Móvil Mes2	Saliente	Flag	OnNet
V10	Flag tráfico Móvil-Móvil Mes3	Saliente	Flag	OnNet
V11	Nro meses tráfico Movil-Móvil	Saliente	Cantidad	OnNet
V12	Promedio Castigado minutos Móvil-Móvil	Saliente	Minuto	OnNet
V13	Promedio llamadas Móvil-Móvil	Saliente	Llamada	OnNet
V14	Promedio duración minutos Móvil-Móvil	Saliente	Minuto	OnNet
V15	Mediana duración minutos Móvil-Móvil	Saliente	Minuto	OnNet
V16	Nro total minutos Móvil-Claro	Saliente	Minuto	OffNet
V17	Nro total llamadas Móvil-Claro	Saliente	Llamada	OffNet
V18	Nro total destinos distintos Móvil-Claro	Saliente	Cantidad	OffNet
V19	Promedio total minutos Móvil-Claro	Saliente	Llamada	OffNet
V20	Mínima duración minutos Móvil-Claro	Saliente	Minuto	OffNet
V21	Máxima duración minutos Móvil-Claro	Saliente	Minuto	OffNet
V22	Flag tráfico Móvil-Claro Mes1	Saliente	Flag	OffNet
V23	Flag tráfico Móvil-Claro Mes2	Saliente	Flag	OffNet
V24	Flag tráfico Móvil-Claro Mes3	Saliente	Flag	OffNet
V25	Nro meses tráfico Movil-Claro	Saliente	Cantidad	OffNet
V26	Promedio Castigado minutos Móvil-Claro	Saliente	Minuto	OffNet
V27	Promedio llamadas Móvil-Claro	Saliente	Llamada	OffNet
V28	Promedio duración minutos Móvil-Claro	Saliente	Minuto	OffNet

V29	Mediana duración minutos Móvil-Claro	Saliente	Minuto	OffNet
V30	Nro total minutos Móvil-Nextel	Saliente	Minuto	OffNet
V31	Nro total llamadas Móvil-Nextel	Saliente	Llamada	OffNet
V32	Nro total destinos distintos Móvil-Nextel	Saliente	Cantidad	OffNet
V33	Promedio total minutos Móvil-Nextel	Saliente	Llamada	OffNet
V34	Mínima duración minutos Móvil-Nextel	Saliente	Minuto	OffNet
V35	Máxima duración minutos Móvil-Nextel	Saliente	Minuto	OffNet
V36	Flag tráfico Móvil-Nextel Mes1	Saliente	Flag	OffNet
V37	Flag tráfico Móvil-Nextel Mes2	Saliente	Flag	OffNet
V38	Flag tráfico Móvil-Nextel Mes3	Saliente	Flag	OffNet
V39	Nro meses tráfico Movil-Nextel	Saliente	Cantidad	OffNet
V40	Promedio Castigado minutos Móvil-Nextel	Saliente	Minuto	OffNet
V41	Promedio llamadas Móvil-Nextel	Saliente	Llamada	OffNet
V42	Promedio duración minutos Móvil-Claro	Saliente	Minuto	OffNet
V43	Mediana duración minutos Móvil-Nextel	Saliente	Minuto	OffNet
V44	Nro total de minutos Otro Ope	Saliente	Minuto	Todas
V45	Nro total de llamadas Otro Ope	Saliente	Llamada	Todas
V46	Nro total destinos distintos Otro Ope	Saliente	Cantidad	Todas
V47	Promedio total minutos Otro Ope	Saliente	Llamada	Todas
V48	Mínima duración minutos Otro Ope	Saliente	Minuto	Todas
V49	Máxima duración minutos Otro Ope	Saliente	Minuto	Todas
V50	Flag tráfico Otro Ope Mes1	Saliente	Flag	Todas
V51	Flag tráfico Otro Ope Mes2	Saliente	Flag	Todas
V52	Flag tráfico Otro Ope Mes3	Saliente	Flag	Todas
V53	Nro meses tráfico Otro Ope	Saliente	Cantidad	Todas
V54	Promedio Castigado minutos Otro Ope	Saliente	Minuto	Todas
V55	Promedio llamadas Otro Ope	Saliente	Llamada	Todas
V56	Promedio duración minutos Otro Ope	Saliente	Minuto	Todas
V57	Mediana duración minutos Otro Ope	Saliente	Minuto	Todas
V58	Nro total de minutos	Saliente	Minuto	Todas
V59	Nro total de llamadas	Saliente	Llamada	Todas
V60	Nro total destinos distintos	Saliente	Cantidad	Todas
V61	Promedio total minutos	Saliente	Llamada	Todas
V62	Mínima duración minutos	Saliente	Minuto	Todas
V63	Máxima duración minutos	Saliente	Minuto	Todas

V64	Flag tráfico Mes1	Saliente	Flag	Todas
V65	Flag tráfico Mes2	Saliente	Flag	Todas
V66	Flag tráfico Mes3	Saliente	Flag	Todas
V67	Nro meses tráfico	Saliente	Cantidad	Todas
V68	Promedio Castigado minutos	Saliente	Minuto	Todas
V69	Promedio llamadas	Saliente	Llamada	Todas
V70	Promedio duración minutos	Saliente	Minuto	Todas
V71	Mediana duración minutos	Saliente	Minuto	Todas
V72	Nro total de sms	Saliente	Cantidad	Todas
V73	Nro total destinos distintos sms	Saliente	Cantidad	Todas
V74	Flag sms Mes1	Saliente	Flag	Todas
V75	Flag sms Mes2	Saliente	Flag	Todas
V76	Flag sms Mes3	Saliente	Flag	Todas
V77	Nro meses sms	Saliente	Cantidad	Todas
V78	Nro de Celdas distintas de Origen de Tráfico	Saliente	Cantidad	Todas
V79	Nro total minutos	Entrante	Minuto	OnNet
V80	Nro total llamadas	Entrante	Llamada	OnNet
V81	Promedio total minutos	Entrante	Llamada	OffNet
V82	Flag tráfico Mes1	Entrante	Flag	OnNet
V83	Flag tráfico Mes2	Entrante	Flag	OnNet
V84	Flag tráfico Mes3	Entrante	Flag	OnNet
V85	Nro meses tráfico	Entrante	Cantidad	OnNet
V86	Promedio Castigado minutos	Entrante	Minuto	OnNet
V87	Promedio llamadas	Entrante	Llamada	OnNet
V88	Promedio duración minutos	Entrante	Minuto	OnNet
V89	Mediana duración minutos	Entrante	Minuto	OnNet
V90	Nro de Celdas distintas de Origen de Tráfico	Entrante	Cantidad	Todas

15 variables derivadas del tráfico:

Cod	Variables	Formula
K1	% Llamadas Móvil-Móvil Salientes	V3/V59
K2	% Minutos Móvil-Móvil Salientes	V2/V58
K3	Dispersión Móvil-Móvil	V4/V3
K4	% Llamadas Claro Saliente	V17/V59

K5	% Minutos Claro Saliente	V16/V58
K6	Dispersión Claro Ope	V18/V17
K7	% Llamadas Nextel Saliente	V31/V59
K8	% Minutos Nextel Saliente	V30/V58
K9	Dispersión Nextel Ope	V32/V31
K10	% Llamadas Otro Ope Saliente	V45/V59
K11	% Minutos Otro Ope Saliente	V44/V58
K12	Dispersión Otro Ope	V46/V45
K13	% Minutos RPM Saliente	V1/V58
K14	% Minutos Salientes sobre el Total	V58/(V58+V79)
K15	% Llamadas Salientes sobre el Total	V59/(V59+V80)

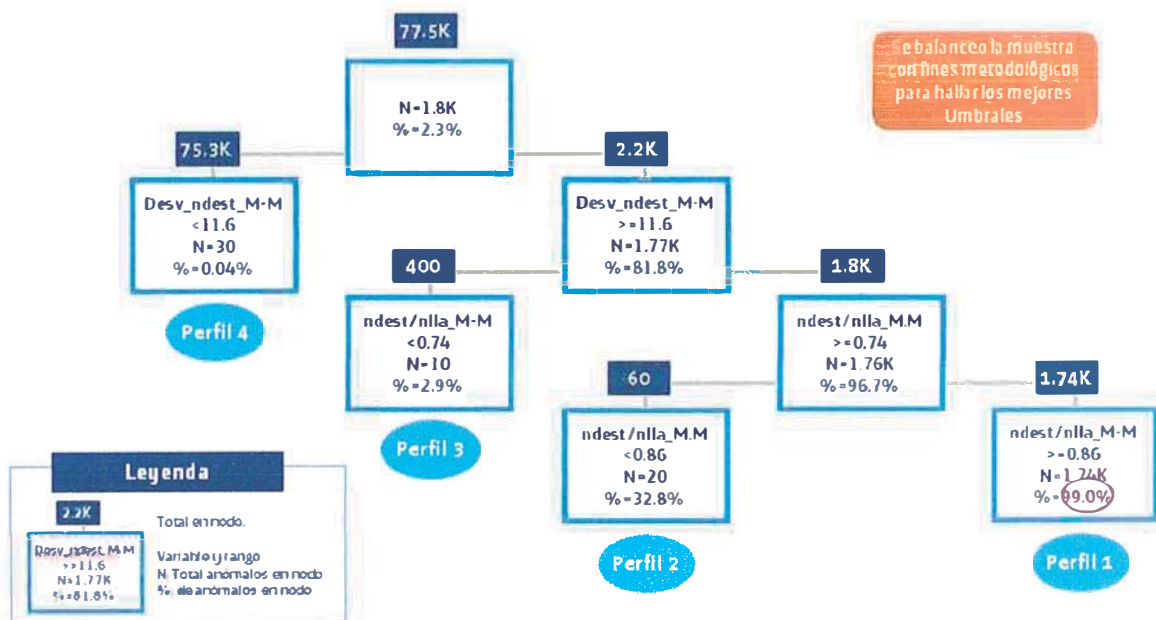
20 variables con información del abonado:

Nro	Variables	Tipo Trafico	Unidad	Tipo Llamada
P1	Fecha de alta	Saliente	Minuto	RPM
P2	Tipo de producto	Saliente	Minuto	OnNet
P3	Modelo de terminal	Saliente	Llamada	OnNet
P4	Ciclo de facturación	Saliente	Cantidad	OnNet
P5	Plan tarifario	Saliente	Llamada	OffNet
P6	Cargo fijo del plan tarifario	Saliente	Minuto	OnNet
P7	Flag RPM	Saliente	Minuto	OnNet
P8	Flag LDI	Saliente	Flag	OnNet
P9	Flag Roaming	Saliente	Flag	OnNet
P10	Tipo cliente	Saliente	Flag	OnNet
P11	Segmento	Saliente	Cantidad	OnNet
P12	Ultima fecha CAEQ	Saliente	Minuto	OnNet
P13	Ultima fecha Cambio de Firma	Saliente	Llamada	OnNet
P14	Ultima fecha de cambio de Plan	Saliente	Minuto	OnNet
P15	Valor Minutos Adicionales	Saliente	Minuto	OnNet
P16	Promedio Facturacion	Saliente	Minuto	OffNet
P17	Ultima fecha cambio limite de crédito	Saliente	Llamada	OffNet
P18	Ultima fecha recarga	Saliente	Cantidad	OffNet
P19	Monto ultima recarga	Saliente	Llamada	OffNet
P20	Ultima fecha migración	Saliente	Minuto	OffNet

4.5. Modelamiento:

Para el modelo se utilizó un algoritmo predictivo de arboles de decisión, por su facilidad en la ejecución, análisis e implementación de los resultados ya que el objetivo final es poder replicar el modelo a toda la base de clientes postpago e identificar aquellos que tienen un perfil de consumo anómalo de tráfico.

Grafico del Árbol de Decisión resultante para el modelo.



4.6. Testeo:

Para este modelo no fue necesario realizar la fase de Testeo puesto que se trata de un modelo de arboles de decisión donde la probabilidad de acierto es muy alta debido a que se está perfilando comportamientos de fraude y usualmente las variables que mas impactan en el objetivo se hacen muy notorias.

CAPÍTULO IV: RESULTADOS

1. RESULTADOS DE LA SOLUCIÓN PLANTEADA

Según los resultados del modelo las variables más importantes y que tienen mayor correlación con el perfil de un cliente fraudulento son:

Nro	Variables	Tipo Llamada
K1	% Llamadas Móvil-Móvil Salientes	V3/V59
K2	% Minutos Móvil-Móvil Salientes	V2/V58
K3	Dispersión Móvil-Móvil	V4/V3
K14	% Minutos Salientes sobre el Total	V58/(V58+V79)
K15	% Llamadas Salientes sobre el Total	V59/(V59+V80)
K10	% Llamadas Otro Ope Saliente	V45/V59
K11	% Minutos Otro Ope Saliente	V44/V58
K12	Dispersión Otro Ope	V46/V45
K13	% Minutos RPM Saliente	V1/V58

Dentro de ellas las 2 variables más relevantes en el modelo son:

- **% Llamadas Móvil-Móvil Salientes:** Según el modelo es altamente probable que aquellos clientes que solo llaman y no reciben llamadas tengan perfil fraudulento..
- **La dispersión de los números de destino M-M:** Esta es la variable que más correlación tiene con el objetivo y que prácticamente identifica a un cliente fraudulento.

CONCLUSIONES Y RECOMENDACIONES

1. CONCLUSIONES

- Los clientes que realizan fraude ByPass Internacional tienen una alta dispersión de sus destinos móvil-móvil (>86%), llaman a muchos números diferentes. Este comportamiento es atípico para un usuario normal de telefonía celular.
- La cantidad de llamadas salientes móvil-móvil sobre el total de llamadas es elevada (>95%). Prácticamente no recibe llamadas.
- Los modelos predictivos de Data Mining aplicados a temas de identificación de perfiles de fraude tienen una alta probabilidad de certeza, para nuestro caso de análisis es de un 96.7%.

2. RECOMENDACIONES

- Por temas de reducción de costos y mayor alcance se recomienda usar el modelo de data mining para la identificación de clientes con perfil de fraude ByPass Internacional; pero no debe usarse como una verdad absoluta para ejecutar cortes de servicio por fraude ya que para ello se necesitan pruebas más objetivas y en este caso se recomienda complementarlo con visitas de campo a los clientes sin previo aviso para verificar el uso de las líneas o realizar llamadas de prueba o satisfacción del cliente con la finalidad de corroborar si responden las llamadas.
- Cada cierto tiempo también sería recomendable usar el sistema IL-Hunter para comprobar algunos hallazgos con el modelo predictivo y así tener mayor sustento para solicitar los cortes al regulador Osiptel.

BIBLIOGRAFÍA

1. Gartner Group. *Data Mining* [en línea]: *SAMBA department*.
<[http://www2.nr.no/documents/samba/research_areas/BAMG/Pattern/Data Mining.html](http://www2.nr.no/documents/samba/research_areas/BAMG/Pattern/Data%20Mining.html)>. [Consulta Noviembre 2012].
2. Chavez, Julio Cesar. *Data Mining Perfiles y Tendencias* [diapositivas].
Lima Perú: Escuela de Postgrado de la UPC, 2012.
3. Kdnuggets. *What main methodology are you using for data mining?*
http://www.kdnuggets.com/polls/2007/data_mining_methodology.html
[Consulta Noviembre 2012].
4. WIKIPEDIA. *Minería de datos* [en línea]:
<http://es.wikipedia.org/wiki/Miner%C3%ADa_de_datos>.
[Consulta Noviembre 2012].
5. P. Naïm, P. Wuillemin, P. Leray, O. Pourret, A. Becker. *Les réseaux bayésiens*, Eyrolles 2004.
6. SAS. *SAS Enterprise Miner* [en línea]:
<<http://www.sas.com/technologies/analytics/datamining/miner/>>.
[Consulta Noviembre 2012].