

**UNIVERSIDAD NACIONAL DE INGENIERIA
FACULTAD DE INGENIERIA INDUSTRIAL Y DE SISTEMAS**



**OPTIMIZACION DE LA DETECCION DE HURTO DE
LINEA EN TELEFONIA FIJA UTILIZANDO LA TECNICA
DE MINERIA DE DATOS**

INFORME DE SUFICIENCIA

**PARA OPTAR POR EL TITULO PROFESIONAL DE:
INGENIERO DE SISTEMAS**

CHRISTIAN JULIO VALDIVIEZO ALTAMIRANO

LIMA, PERU

2012

DEDICATORIA

Especialmente a mi Madre, mis hermanos Juan y Cora,
además de todas aquellas personas que me brindaron
su apoyo y confianza para que logre esta meta,
a mi Padre que partió hacia el Señor antes.

INDICE

INFORME DE SUFICIENCIA	1
DEDICATORIA.....	2
INDICE	3
DESCRIPTORES TEMATICOS.....	6
RESUMEN EJECUTIVO	7
INTRODUCCION	9
CAPÍTULO I: PENSAMIENTO ESTRATÉGICO	11
1.1. DIAGNÓSTICO FUNCIONAL.....	11
1.1.1. HISTORIA.....	12
1.1.2. ORGANIZACIÓN	13
1.1.3. SERVICIOS	14
1.1.4. CLIENTES.....	19
1.1.5. PROVEEDORES.....	21
1.1.6. PROCESOS.....	22
1.2. DIAGNÓSTICO ESTRATÉGICO.....	23
1.2.1. VISIÓN.....	23
1.2.2. MISIÓN.....	24
1.2.3. VALORES	24
1.2.4. ANÁLISIS INTERNO	25
1.2.5. ANÁLISIS EXTERNO	26
1.2.6. OBJETIVOS ESTRATEGICOS	28
1.2.7. MATRIZ FODA	29
CAPÍTULO II: MARCO TEÓRICO Y METODOLÓGICO	31
2.1. FRAUDE EN LAS TELECOMUNICACIONES	31

2.1.1.	ANTECEDENTES.....	31
2.1.2.	FRAUDE EN LAS TELECOMUNICACIONES	32
2.1.3.	PRINCIPALES TIPOLOGÍAS DE FRAUDE EN TELECOMUNICACIONES.....	32
2.1.4.	EL PERFIL DEL HURTADOR DE LINEA.....	39
2.2.	JEFATURA DE LUCHA CONTRA EL FRAUDE – SUPERVISION RED FIJA.....	45
2.2.1.	SERVICIO DE CONTROL DE FRAUDE.....	46
2.3.	MINERIA DE DATOS	50
2.3.1.	FASE DE UN PROCESO CLASICO DE MINERIA DE DATOS	51
2.3.2.	METODOLOGÍAS DE APLICACIÓN DE LA MINERIA DE DATOS	65
2.3.3.	APLICACIÓN DE LA MINERIA DE DATOS Y TENDENCIAS.....	74
2.3.4.	DIFICULTADES EN LA IMPLEMENTACION DE LA MINERIA DE DATOS	76
CAPÍTULO III: PROCESO DE TOMA DE DECISIONES		78
3.1.	OBJETIVO	78
3.1.1.	OBJETIVO GENERAL	79
3.1.2.	OBJETIVO ESPECIFICO.....	79
3.2.	IDENTIFICACIÓN DE PROBLEMA.....	80
3.3.	PLANTEAMIENTO DE ALTERNATIVAS DE SOLUCIÓN.....	83
3.3.1.	ALTERNATIVA 1: MEJORA DEL PROCESO DE GENERACIÓN DE ALERTAS DE HURTO DE LÍNEA.....	83
3.3.2.	ALTERNATIVA 2: MODELAMIENTO DEL HURTO DE LÍNEA MEDIANTE LA TÉCNICA DE MINERIA DE DATOS.....	84
3.4.	SELECCIÓN DE UNA ALTERNATIVA DE SOLUCIÓN.....	85
3.4.1.	CUADRO DE EVALUACION DE CRITERIOS VS ALTERNATIVAS.....	86
3.5.	PLANES DE ACCIÓN PARA DESARROLLAR LA SOLUCIÓN ELEGIDA.....	87
3.5.1.	FASE I: COMPRENSIÓN DEL NEGOCIO	90
3.5.2.	FASE II: COMPRENSIÓN DE LOS DATOS.....	103
3.5.3.	FASE III: PREPARACIÓN DE LOS DATOS.....	106
3.5.4.	FASE IV: MODELADO.....	109
3.5.5.	FASE V: EVALUACIÓN.....	110
3.5.6.	FASE VI: IMPLEMENTACIÓN	112
CAPITULO IV: ANÁLISIS DE RESULTADOS.....		114
4.1.	SELECCIÓN DE CRITERIOS DE EVALUACIÓN.....	114

4.2. INFORMACIÓN DE SITUACIÓN ACTUAL.....	115
4.3. RESULTADOS DE LA SOLUCIÓN PLANTEADA.....	120
CONCLUSIONES Y RECOMENDACIONES.....	123
CONCLUSIONES.....	123
RECOMENDACIONES.....	124
GLOSARIO DE TERMINOS.....	126
BIBLIOGRAFIA.....	131
ANEXOS.....	133

DESCRIPTORES TEMATICOS

Hurto de Línea

Fraude por Suscripción

Empresa de Telecomunicaciones

Averías

Clip-On

Fraude Técnico

Minería de datos

DataWareHouse

Redes neuronales

Tipología

DataMining

RESUMEN EJECUTIVO

El fraude en las telecomunicaciones es una actividad que vulnera las barreras de entrada o los medios de control con que cuentan las empresas de telecomunicaciones. Actualmente el fraude que presenta mayor impacto es el hurto de línea debido a que afecta la calidad del servicio porque vulnera la planta externa, incrementa la cantidad de reclamos presentados por desconocimiento del tráfico facturado, incrementa el monto de pérdida porque al ser tráfico no realizado por el cliente principalmente a países de riesgo se termina ajustando la facturación, además termina impactando en la imagen de la empresa por brindar un servicio de mala calidad o vulnerable. Se necesita principalmente mejorar la generación de alertas para la detección oportuna del hurto de línea para lo cual se utilizará la técnica de minería de datos basado en la metodología CRISP-DM.

Empezamos realizando un análisis funcional y estratégico de **tgestiona** que es una empresa especializada en brindar soluciones de negocio confiables, eficientes y oportunas, a través de la gestión de servicios de apoyo a la gestión administrativa, bajo un modelo de tercerización, creando valor agregado a las empresas. **tgestiona** brinda el servicio de Control de Fraude a la empresa de telecomunicaciones más grande del Perú.

Se ha desarrollado un marco teórico buscando abarcar todos los tópicos necesarios para entender los conceptos de fraude y sus tipologías, además cómo la Supervisión de Lucha contra el Fraude busca principalmente reducir

el impacto económico que genera el fraude en las empresas de telecomunicaciones. Se desarrolla la técnica de minería de datos sus principales etapas de un proyecto de minería de datos, las metodologías para una correcta implementación y sus principales usos.

A continuación se describe el objetivo principal del informe, donde se busca encontrar la solución que presente mayor beneficio al problema que se ha detectado, que es reducir el impacto económico y de imagen debido al fraude por hurto de línea que afecta a la principal empresa de telecomunicaciones del Perú. Se han planteado dos alternativas de solución, se describen cada una de ellas, para luego realizar una evaluación de las mismas por los criterios principales que fueron elegidos por un equipo de expertos de la Lucha contra el Fraude. Luego se describe toda la implementación realizada para desarrollar la alternativa que obtuvo mayor puntaje en la evaluación anterior.

Finalmente desarrollamos el análisis económico de la solución implementada, seleccionando los principales indicadores y criterios de evaluación de proyectos. Se puntualizan las conclusiones y recomendaciones que se han encontrado a lo largo del desarrollo e implementación de la solución.

INTRODUCCION

Desde la década de los noventa, cuando las telecomunicaciones en el Perú se incrementaron raudamente, debido a la privatización, se implementaron nuevas tecnologías, infraestructuras y servicios, pero simultáneamente también se diseñaron y evolucionaron nuevos métodos de fraude que buscaban aprovecharse de las facilidades comerciales y vulnerabilidades tecnológicas que brindaban las empresas de telecomunicaciones, logrando así que la calidad del servicio e infraestructura se vean afectadas.

Tomando en cuenta el contexto antes presentado y debido a que el Fraude por Suscripción (Suplantación de la Identidad) ha disminuido por la implementación de barreras de entrada en el proceso de venta y métodos de detección preventiva, además se continua monitoreando el consumo anómalo en los diversos servicios telefónicos fijos, por estas razones los fraudulentos han migrado sus tácticas para vulnerar las facilidades de la planta instalada y realizar hurto de línea afectando a las empresas de telecomunicaciones. Se desarrollará para la telefonía fija un modelo utilizando la técnica de la Minería de Datos para la optimización de la generación de alertas para la detección del Hurto de Línea. El modelo propuesto buscará identificar los casos que presenten perfil de fraude por hurto de línea con una alta probabilidad de acierto y de esta manera lograr reducir el impacto económico en las pérdidas de ajustes por fraude, disminuir el número de reclamos de los clientes que desconocen el tráfico facturado y reducir la manipulación de las facilidades técnicas por terceros que afecta directamente la calidad del servicio.

Una postura firme contra el fraude en las telecomunicaciones no solo contribuirá a reducir los costos y precios para el usuario final, sino que las empresas puedan brindar servicios de calidad e implementar nuevas tecnologías para desarrollar el estilo de vida de los usuarios.

CAPÍTULO I

PENSAMIENTO ESTRATÉGICO

1.1. DIAGNÓSTICO FUNCIONAL

tgestiona es una empresa internacional especializada en brindar soluciones de negocio confiables, eficientes y oportunas, a través de la gestión de servicios de apoyo a la gestión administrativa, bajo un modelo de tercerización, creando valor agregado al negocio.

Tiene presencia en cuatro países: España, Brasil, Argentina y Perú. Cuenta con más de 10 años de experiencia en la prestación de servicios a empresas de diferentes sectores económicos, como telecomunicaciones, banca, industria, servicios, educación, entre otros.

La Dirección de Seguridad y Protección cuenta con el know how que le avala en los procedimientos de gestión en seguridad y lucha contra el fraude para el bienestar de sus clientes. Al saber la importancia de unir la seguridad a la cadena de valor, alineándose con los objetivos empresariales al menor tiempo y costo posible, agregando valor y con un retorno de inversión razonable. Es por esta razón que cuenta con personal y soporte tecnológico altamente calificado en materia de soluciones integrales en seguridad empresarial.

1.1.1. HISTORIA

[TGES01] Hacia mediados del año 2000 en el Grupo Telefónica nació un nuevo desafío; El desafío de poder contar con el mejor Centro de Servicios Compartidos para todo el Grupo, con compromiso de eficiencia, calidad y especialización en los servicios brindados. Ese fue el nacimiento de **tgestiona**, desde ese momento, dirigió las soluciones de negocio hacia la satisfacción de las necesidades de los clientes del Grupo Telefónica y al desarrollo de su personal, manteniendo el enfoque en la optimización constante de sus servicios y procesos.

En el 2003, durante los primeros años, la misión tuvo como objetivo fundamental lograr la optimización de procesos administrativos para lograr sinergias y generar eficiencias para cada una de las empresa del Grupo, de tal forma que cada una de ellas lograra una ventaja competitiva en sus respectivos sectores económicos; orientándonos hacia la capacitación y especialización de nuestro equipo, obteniendo certificaciones internacionales en gestión de calidad y forjando nuestra cultura de vocación de servicio.

En el 2006, el tiempo, el know how y el compromiso de cada uno de nuestros colaboradores, hizo que poco a poco, **tgestiona** logrará la experiencia y efectividad necesarias en cada una de sus soluciones de negocio ofrecidas. Orientando los esfuerzos en generar valor a los servicios de los clientes, adecuando la oferta a las exigencias del mercado externo, se inicio de esta forma, un modelo de fidelización de clientes, que abarca desde la cultura de servicio del personal hasta la fidelización en sí del cliente.

Desde el 2011, **tgestiona** es hoy sinónimo de innovación y evolución permanente. La nueva economía mundial exige cada vez más a las empresas a focalizar y fortalecer al máximo las especialidades y competencias organizacionales, buscando constantemente mayores eficiencias. Es allí donde halla valor su labor. Porque para aprovechar las oportunidades en un mercado competitivo, no existe mejor alternativa que contar con la asesoría de expertos en soluciones de negocio. En **tgestiona** reafirman su compromiso con los clientes al convertirse en sus socios estratégicos y ser capaces de brindarles soluciones a nivel internacional con la garantía de contar siempre con servicios de calidad que reflejen la aplicación de las mejores prácticas del mercado.

1.1.2. ORGANIZACIÓN

tgestiona presenta una organización jerárquica que a continuación se describe:

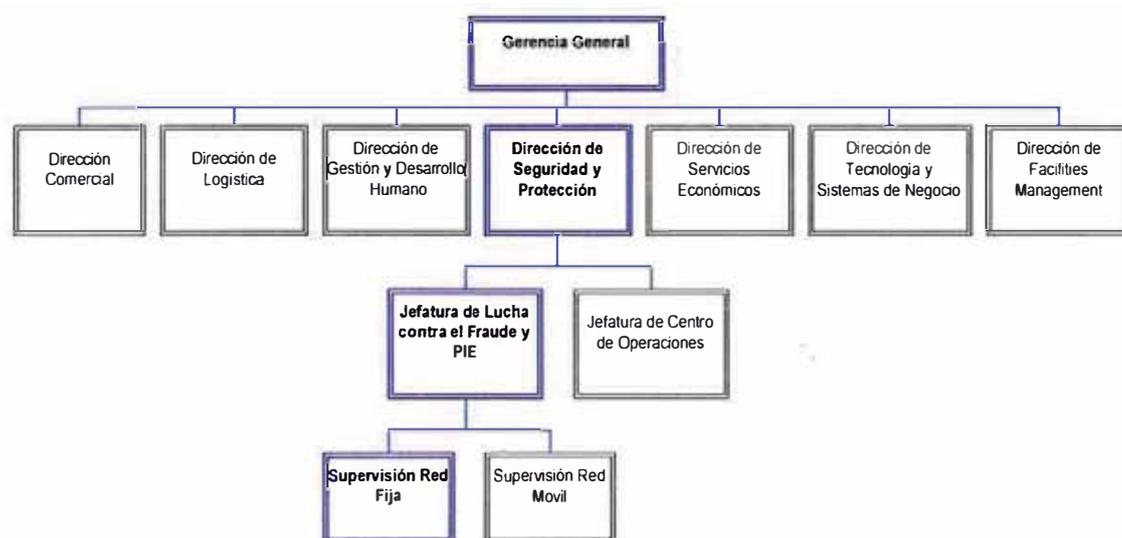


Gráfico 1.1: Organigrama de **tgestiona** (Fuente: **tgestiona**,
Elaboración: Propia)

1.1.3. SERVICIOS

[TGES02] *tgestiona* es una empresa que brinda soluciones de negocio o soluciones integrales empresariales, a continuación se describen las principales.



Gráfico 1.2: Servicios de *tgestiona* (Fuente y Elaboración: *tgestiona*)

1.1.3.1. Principales servicio de Tgestiona

1.1.3.1.1. *Tecnología y Sistemas de Negocio*

- Consultoría de Sistemas
- Mejora de Procesos
- Testing y Quality Assurance de aplicaciones

- Software Factory
- Aseguramiento de la Calidad
- Consultoría de Base de Datos
- Soluciones y Soporte SAP
- Soluciones de Gestión Documental
- Consultoría de Seguridad
- Gestión de PMO
- Soluciones y Soporte de Plataformas de Recursos Humanos

1.1.3.1.2. Gestión Humana

- Búsqueda de Talentos
- Proyección Social
- e-Learning
- Gestión de Nómina y Beneficios

1.1.3.1.3. Consultoría y Soluciones de Negocio

- Soluciones para la Optimización de Proceso
- Optimización de Modelos
- Gestión de Cambio
- Soluciones para la Mejora en la Calidad de Procesos
- Soluciones de Análisis Organizacional
- Integración de Soluciones y ERPs

1.1.3.1.4. Gestión Logística

- Gerencia de Operaciones Logísticas
- Servicio de Planificación y Aprovisionamiento

1.1.3.1.5. Contabilidad, Tributos y Pagos

- Contabilidad Analítica
- Conciliación Bancaria
- Gestión de Activos Fijos y Existencias
- Contabilidad Regulatoria
- Gestión de Cajas y Bancos
- Gestión del Cumplimiento Tributario
- Administración de Valores
- Gestión de Ventanilla Única

1.1.3.1.6. Gestión de Cobranzas

- Gestión de Cobranzas
- Registro y Control
- Soporte Administrativo y Control de Recaudación

1.1.3.1.7. Facilities Management

- Gestión de Espacios
- Desarrollo de Proyectos de Ingeniería y Arquitectura
- Mantenimiento Integral de Edificios
- Gestión y Administración Patrimonial
- Gestión de Servicios Generales

1.1.3.2. Principales servicios de la Dirección de Seguridad y Protección

A continuación se describen con mayor detalle los servicios ofrecidos por la Dirección de Seguridad y Protección:

1.1.3.2.1. Gestión de Seguridad Integral

El enfoque que ofrece es una gestión rentable de los riesgos de seguridad basada en actividades orientadas a la identificación, detección y evaluación de riesgos, a lo largo de la cadena de valor, para gestionarlos en el marco de un modelo de agregación de valor, de aporte de beneficios y de contribución a resultados. Los riesgos de seguridad se manifiestan en diferentes formas tales como incendios, robos, asaltos, desastres naturales, etc., contra personas, activos, recursos, actividades y objetivos de su empresa.

Se cuenta con los siguientes subservicios:

- Seguridad Humana
- Seguridad Electrónica
- Monitoreo de alarmas

1.1.3.2.2. Consultoría de Seguridad Integral

Elabora propuestas de valor para el logro de una seguridad rentable. Evalúa la mejor combinación de medidas de seguridad que asegure un equilibrio de costo/beneficio razonable. Alinea la seguridad con los objetivos de la Empresa con proyecciones de implementación de corto, mediano y largo plazo que se adecúen a la disponibilidad de recursos.

Considera que el análisis preventivo de riesgos, analizados en términos de probabilidad, vulnerabilidad y consecuencias, es vital para la seguridad de la empresa. Los resultados permiten conocer toda la problemática que debe considerar tanto en

relación a situaciones de emergencia, como para la gestión rentable de riesgos. La idea central es proponer la incorporación del ciclo de vida de seguridad en la cadena de valor y en el ciclo de vida del negocio transformando a los servicios de seguridad en aliados y en una ventaja competitiva de la Empresa.

Se cuenta con los siguientes subservicios:

- Estudios de Seguridad
- Planes de Emergencia
- Seguridad Humana
- Concientización en Seguridad
- Seguridad Electrónica
- Monitoreo de Alarmas
- Capacitación en Seguridad Empresarial
- Elaboración de Especificaciones Técnicas

1.1.3.2.3. Consultoría en Riesgos y Fraude

Mejora su competitividad en el mercado con un modelo rentable y escalable. Agrega valor, contribuye a sus resultados y transfiere beneficios en forma de ahorros e ingresos, reducciones de pérdidas, incrementos de productividad y mejores tiempos de respuestas ante modalidades de fraude. Realiza actividades que permitan analizar e identificar problemas relacionados ante la eventualidad de la ocurrencia de un fraude, perpetrado por personal interno, malos clientes, organizaciones o personas fraudulentas en los diferentes servicios que el negocio ofrezca, minimizando el impacto en los aspectos económicos y financieros.

Se cuenta con los siguientes subservicios:

- Monitoreo e identificación de Transacciones
- Análisis de Perfiles de Consumo para la Generación de alertas con Patrones de Anómalos
- Análisis de Riesgos de Productos, Servicios y Procesos
- Evaluación de Expedientes de Suscriptores

1.1.4. CLIENTES

[TGES03] *tgestiona* al ser una empresa que brinda servicios variados cuenta con un diverso grupo de clientes, que le puede brindar los servicios que ofrece, sin embargo se pueden segmentar los clientes de la siguiente manera:



Gráfico 1.3: Principales grupos de clientes de *tgestiona* (Fuente y Elaboración: *tgestiona*)

- Grupo Telefónica: Principal Cliente y razón de ser primigenia de **tgestiona**, conformado por diversas empresas dedicadas principalmente a brindar servicios de telecomunicaciones.
- Sector Financiero: Empresas conformadas por los Bancos, Cajas de Ahorro y afines relacionadas al sector financiero.
- Sector Gobierno: Entidades de los diversos sectores productivos o económicos, que son supervisados o administrados por el Estado.
- Servicios Públicos: Empresas que brindan principalmente la generación o distribución de servicios públicos como agua, desagüe, alcantarillado y generación eléctrica.
- Sector Minería e Industria: Empresas que se desempeñan en el rubro minería y la producción industrial.
- Servicios Varios: Se considera a todas las demás empresas que no son pertenecen a las segmentación descritas anteriormente.

1.1.4.1. Clientes de **tgestiona**

Dentro de los principales Clientes que cuenta **tgestiona** están:

- Grupo Telefónica: Movistar, Cable Mágico, Telefónica, Atento, Terra, Media Networks, Fundación Telefónica, Cable Mágico, etc.
- Sector Financiero: Banco de Crédito, Interbank, Scotiabank, BBVA Continental, Mi Banco.
- Sector Gobierno: Fondo Mi Vivienda, ONP, Marina de Guerra, Banco Central de Reserva, ENAPU, PeruPetro, ONPE, CORPAC.
- Servicios Varios: PUCP, UPC, Universidad Científica del Sur, Boticas BTL, Casa Andina, DHL, etc.
- Servicios Públicos: Edelnor, Luz del Sur, Sedapal y ENOSA.

- Sector Minería e Industria: Rio Alto Mining Limited, Hochshield, Indra, Lindley, Telemark, Digitex, TASA, etc.

1.1.4.2. Clientes de la Dirección de Seguridad y Protección

Dentro de los principales Clientes que cuenta la Dirección de Seguridad y protección están:

- Grupo Telefónica: Movistar, Cable Mágico, Telefónica, Atento, Terra, Media Networks, Cable Mágico, etc.
- Sector Financiero: Banco de Crédito, Mi Banco.
- Sector Gobierno: Banco Central de Reserva, CORPAC.
- Servicios Varios: CARSA, IPSOS Apoyo.
- Sector Minería e Industria: AIB Agroindustria, Hochshield.

1.1.5. PROVEEDORES

tgestiona al brindar una gama diversa de servicios, cuenta con un variado grupo de proveedores, entre los que se segmentan de la siguiente manera:

- Trabajos de edificación y acondicionamiento de propiedades.
- Mudanzas y mantenimiento integral de inmuebles.
- Transporte de personas, vehículos en alquiler o renting.
- Mantenimiento de vehículos.
- Aprovisionamiento, almacenaje y manipulación.
- Transporte y distribución diversa.
- Servicios de seguridad física.
- Servicios técnicos de Ingeniería de sistemas.
- Hardware microinformática, PC's y portátiles.
- Almacenamiento y mantenimiento sistemas o productos SW.
- Agencias de viajes.

- Material de oficina y consumibles informática.
- Producción gráfica marketing, promocional, eventos y punto de venta.

1.1.6. PROCESOS

En *tgestiona* se cuenta con un proceso general para los servicios que brindan las diferentes direcciones que se describe a continuación:

1.1.6.1. Mapa de procesos de *tgestiona*

En los procesos de *tgestiona*, el principal proceso es la gestión de servicios, donde cada Dirección desarrolla sus principales servicios.

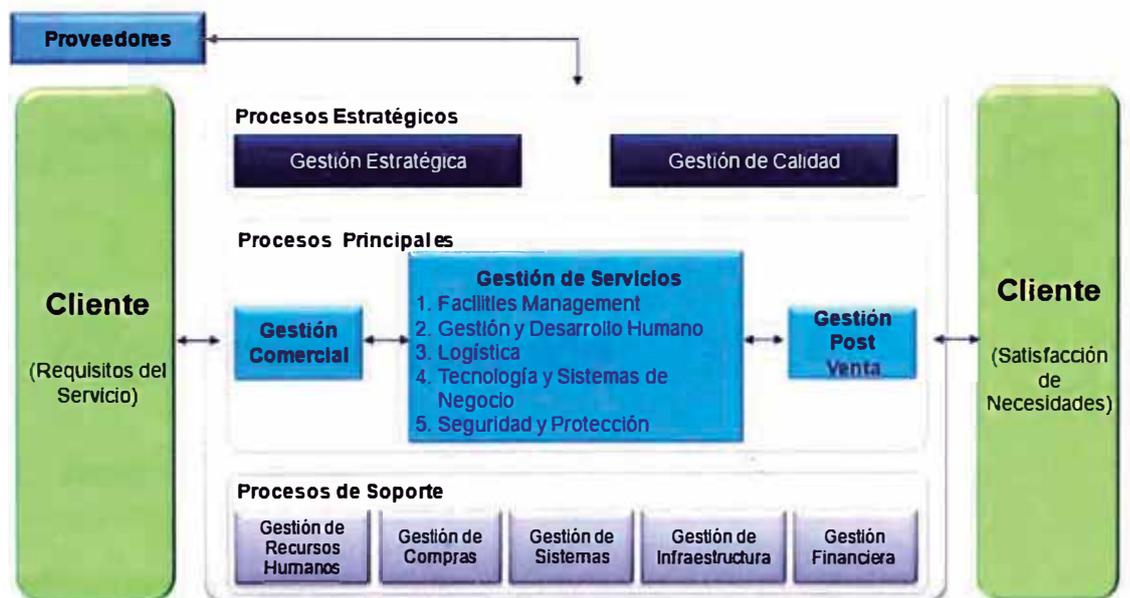


Gráfico 1.4: Mapa de procesos de *tgestiona* (Fuente y Elaboración: *tgestiona*)

1.1.6.2. Mapa de procesos de la Dirección de Seguridad y Protección

El servicio principal de la Dirección de Seguridad y Protección es la detección de fraude, se describen a continuación la totalidad de sus procesos:

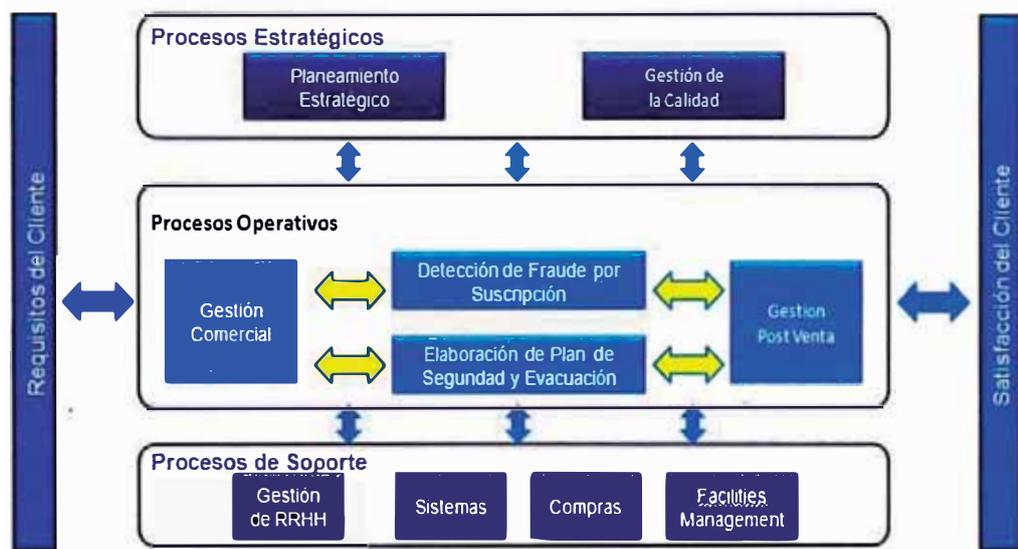


Gráfico 1.5: Mapa de procesos de Dirección de Seguridad y Protección (Fuente y Elaboración: *tgestiona*)

1.2. DIAGNÓSTICO ESTRATÉGICO

1.2.1. VISIÓN

[TGES04] "Ser la organización líder en la prestación de servicios de soluciones de negocio a nivel nacional e internacional, reconocida por contribuir al crecimiento sostenido y rentable de nuestros clientes y de nuestro negocio, y al desarrollo de nuestro equipo humano y de la sociedad."

1.2.2. MISIÓN

[TGES04] "Somos una organización internacional especializada en brindar servicios de soluciones de negocio confiables, eficientes y oportunas, para permitir a nuestros clientes enfocarse en las actividades que les generen mayor valor. Lo hacemos con un equipo humano comprometido y capacitado, asegurando un crecimiento sostenido y rentable para nuestros accionistas y contribuyendo al desarrollo de nuestros empleados y de la sociedad."

1.2.3. VALORES

[TGES04] Las políticas están alineadas con los valores definidos para la empresa, que modelan el actuar de cada una de las personas que conforman *tgestiona* y permite generar y mantener la confianza mediante su ejercicio, fundamental para la construcción de equipos de trabajo de alto performance. Nuestros valores son los siguientes:

- Imparcialidad
- Credibilidad
- Respeto
- Innovación
- Confianza



Gráfico 1.6: Valores de *tgestiona* (Fuente y Elaboración: *tgestiona*)

1.2.4. ANALISIS INTERNO

1.2.4.1. Fortalezas

- Experiencia y Know How en la prestación de servicios de soporte.
- Procesos Certificados con ISO 9001:2000, CMMI, PMI, entre otros.
- Tecnología de punta y vanguardia. Posee plataformas informáticas para manejo de información con soporte tecnológico de primer nivel (SAP, Meta4, IXOS).
- Personal capacitado, calificado y experto en las actividades que realiza.
- Ser parte del Grupo Telefónica, que cuenta con presencia regional y brinda un respaldo Financiero.
- Capacidad instalada a nivel nacional.
- Primera empresa de servicios en el País a nivel de facturación.
- Experiencia en la identificación de riesgos en telecomunicaciones.
- Experiencia en el análisis transaccional de altos volúmenes.
- Economías de escala.

1.2.4.2. Debilidades

- Los sistemas estándares del Grupo Telefónica no necesariamente son la mejor herramienta para competir en el mercado.
- Falta de gestión por indicadores en algunos procesos.
- Imagen de Telefónica no ayuda a la de **tgestiona**, percepción por parte de los clientes empresariales que el mal servicio de Telefónica es extrapolable a **tgestiona**.

- Los procesos establecidos en el Grupo Telefónica no permiten actuar con la rapidez que el mercado demanda.
- Pocas referencias y experiencia en prestación de servicios al mercado externo.
- **tgestióna** no es conocida en el mercado como empresa prestadora de servicios o sino solo como una empresa que brinda de servicios de telecomunicaciones.
- Falta de una estrategia comercial.
- Altos Costos corporativos sobre los que no existe margen de maniobra, y que impacta en la competitividad de los precios.
- Dificultad para reducir y variabilizar el costo de la nómina.
- Capacidad instalada al tope en servicios como Logística y SSEE.
- Falta de alianzas estratégicas con proveedores de servicios.
- No se cuenta con un Sistema Antifraude.
- Rigidez de las políticas y recursos limitados para los desarrollos de sistemas in house.
- Acceso limitado a la información requerida para la prestación del servicio.
- Rigidez en autorización del uso de programas no corporativos.
- Limitado poder de autoridad para casos de fraude en el negocio.
- Falta de identificación del personal.
- Alta rotación de personal.

1.2.5. ANÁLISIS EXTERNO

1.2.5.1. Oportunidades

- Incursión y acogida en el mercado externo por contar con el respaldo del Grupo Telefónica.

- Incremento en la participación del mercado debido al crecimiento de la tercerización de servicios.
- Crecimiento económico en el país y la región y entorno macroeconómico estable.
- Incipiente mercado de servicios integrales en el entorno local.
- Tendencias a la exportación de servicios y ventas a nivel internacional.
- Nuevas tendencias de alianzas con proveedores.
- Tendencia a la disminución de los costos en las empresas.
- No existe competencia en el mercado local que brinde la totalidad de los servicios ofrecidos.
- Servicio a empresas con elevado número de transacciones financieras por fraude externo o transacciones comerciales por fraude interno.
- Problemas técnicos y comerciales en soporte Servicio de atención al Cliente.

1.2.5.2. Amenazas

- Potencial Internalización de actividades o tercerización con otros proveedores por parte de clientes del Grupo Telefónica.
- Imposibilidad de eliminar costos fijos en **tggestiona** principalmente de nómina debido a convenios, ante la eventual internalización o pérdida de servicios al Grupo Telefónica:
- Incremento de competidores en nichos donde participamos, prestando servicios independientes especializados, pero aún no integrales.
- Aparición de nuevas empresas de servicios y guerra de precios entre los actuales competidores.
- Leyes laborales que hacen menos flexible la contratación de personal - Rigidez.

- Tendencias del Grupo Telefónica a no reconocer incremento de tarifas por inflación.
- Bajo tamaño del mercado en el país.
- Creación de unidades del servicio en el Grupo Telefónica que generan duplicidad de funciones y dilación en las gestiones de fraude
- Cambios en las legislaciones vigentes sobre tercerización de actividades.
- Crisis Mundial

1.2.6. OBJETIVOS ESTRATEGICOS

Las principales estrategias de negocio se enfocan en los siguientes cuatro ejes estratégicos, Clientes, Procesos, Personas y Financiero, la siguiente gráfica muestra los objetivos de cada uno de los ejes.



Gráfico 1.7: Objetivos estratégicos de *tgestiona* (Fuente y Elaboración: *tgestiona*)

1.2.7. MATRIZ FODA

Ya hemos descrito las Oportunidades y Amenazas, las Debilidades y Fortalezas, a continuación describimos los Objetivos Estratégicos y las estrategias a mediano y largo plazo. En la siguiente matriz se observan cómo se encuentran relacionados los términos antes descritos.

	Análisis Externo	OPORTUNIDADES	AMENAZAS
Análisis Interno		<ul style="list-style-type: none"> • Incursión y acogida en el mercado externo por contar con el respaldo del Grupo Telefónica. • Incremento en la participación del mercado debido al crecimiento de la tercerización de servicios. • Crecimiento económico en el país y la región y entorno macroeconómico estable. • Incipiente mercado de servicios integrales en el entorno local. • Tendencias a la exportación de servicios y ventas a nivel internacional. • Nuevas tendencias de alianzas con proveedores. • Tendencia a la disminución de los costos en las empresas. • No existe competencia en el mercado local que brinde la totalidad de los servicios ofrecidos. • Servicio a empresas con elevado número de transacciones financieras por fraude externo o transacciones comerciales por fraude interno. • Problemas técnicos y comerciales en soporte Servicio de atención al Cliente. 	<ul style="list-style-type: none"> • Potencial internalización de actividades o tercerización con otros proveedores por parte de clientes del Grupo Telefónica. • Imposibilidad de eliminar costos fijos en tgestiona. • Incremento de competidores en nichos donde participamos. • Aparición de nuevas empresas de servicios y guerra de precios entre los actuales competidores. • Leyes laborales que hacen menos flexible la contratación de personal - Rigidez. • Tendencias del Grupo Telefónica a no reconocer incremento de tarifas por inflación. • Bajo tamaño del mercado en el país. • Creación de unidades del servicio en el Grupo Telefónica que generan duplicidad de funciones y dilación en las gestiones de fraude. • Cambios en las legislaciones vigentes sobre tercerización de actividades. • Crisis Mundial
FORTALEZAS	<ul style="list-style-type: none"> • Experiencia y Know How en la prestación de servicios de soporte. • Procesos Certificados con ISO 9001:2000, CIMIL, PMI, etc. • Tecnología de punta y vanguardia. • Personal capacitado, calificado y experto en las actividades que realiza. • Ser parte del Grupo Telefónica. • Capacidad instalada a nivel nacional. • Primera empresa de servicios en el País a nivel facturación. • Experiencia en la identificación de riesgos en telecomunicaciones. • Experiencia en el análisis transaccional de altos volúmenes. • Economías de escala. 	<ul style="list-style-type: none"> • Programa de Gestión del Conocimiento. • Desarrollar Certificaciones Internacionales en nuestros procesos. • Programa de Retención de Talentos. • Capitalizar oportunidades de ventas en Brasil, España y Chile. • Desarrollo de nuevos productos y servicios de alta demanda en el mercado. • Expansión en el mercado interno Perú (Nuevos clientes externos al Grupo Telefónica). • Expansión del mercado externo Regional (Nuevos clientes del Grupo Telefónica fuera de Perú) • Posicionamiento en el Sector Público. 	<ul style="list-style-type: none"> • Programa de Eficiencias Operativas. • Programa de Branding (Posicionamiento de la marca "tgestiona"). • Implementación del Sistema de Calidad.
DEBILIDADES	<ul style="list-style-type: none"> • Los sistemas estándares del Grupo Telefónica no necesariamente son la mejor herramienta para competir en el mercado. • Falta de gestión por indicadores en algunos procesos. • Imagen de Telefónica no ayuda a la de tgestiona. • Los procesos establecidos en el Grupo Telefónica no permiten actuar con la rapidez que el mercado demanda. • Pocas referencias y experiencia en prestación de servicios al mercado externo. • Tgestiona no es conocida en el mercado como empresa prestadora de servicios o sino solo como una empresa que brinda de servicios de telecomunicaciones. • Falta de una estrategia comercial. • Altos Costos corporativos sobre los que no existe margen de maniobra. • Dificultad para reducir y variabilizar el costo de la nómina. • Capacidad instalada al tope en servicios como Logística y SSEE. • Falta de alianzas estratégicas con proveedores de servicios. • No se se cuenta con un Sistema Antifraude. • Rigidez de las políticas y recursos limitados para los desarrollos de sistemas in house. • Acceso limitado a la información requerida para la prestación del servicio. • Rigidez en autorización del uso de programas no corporativos. • Limitado poder de autoridad para casos de fraude en el negocio. • Falta de identificación del personal. • Alta rotación de personal. 	<ul style="list-style-type: none"> • Explotar alianza con Segmento Empresas de Telefónica • Establecer Programas Motivacionales. • Fomentar el desarrollo del Clima Laboral. • Simplificar y acelerar los procesos de contratación de personal, administrativos y de aprovisionamiento. • Ventas al mercado externo de servicios económicos, organización y seguridad. • Innovar en la ventas de nuestros servicios. • Identificación de proveedores externos para formar alianzas estratégicas. • Desarrollar el capital humano, integrado, fidelizado y capacitado. • Fortalecer la cultura de servicio orientada al Cliente. • Fomentar una cultura de creatividad e innovación. • Incrementar el nivel de acceso a la información del cliente. 	<ul style="list-style-type: none"> • Establecer Plan Comercial. • Establecer Programa de Mantenimiento y Fidelización de clientes. • Acelerar y simplificar nuestros procesos. • Focalización en Sectores Claves. • Desarrollar nuevos canales de venta • Eliminación de servicios cuya rentabilidad y calidad no sean sostenibles en el tiempo. • Posicionamiento de tgestiona como proveedor de servicios empresariales.

Cuadro 1.1: Matriz FODA de **tgestiona** (Fuente: **tgestiona**, Elaboración: Propia)

CAPÍTULO II

MARCO TEÓRICO Y METODOLÓGICO

2.1. FRAUDE EN LAS TELECOMUNICACIONES .

2.1.1. ANTECEDENTES

Durante más de ochenta años el sector de telecomunicaciones fue desarrollado casi exclusivamente por inversores extranjeros, hasta que en 1969 se creó la empresa estatal ENTEL PERU y la Compañía Peruana de Teléfonos (CPT) para proporcionar los servicios de llamadas locales, nacionales e internacionales. La llegada al poder del presidente Fujimori en 1990, el sector de telecomunicaciones se privatizó por completo. La promoción de la inversión privada en este sector se inició en 1991. En el año 1994, se promulga la Ley de Telecomunicaciones que permitía la inversión privada y la libre competencia. En aquel mismo año, el Estado vende aquellas dos empresas de telecomunicaciones a Telefónica de España S.A. a partir de entonces, se da principio al proceso de expansión, con una mayor penetración a estratos medios y bajos social y económicamente [DEPE01]. Al brindar mayores facilidades para tramitar un servicio telefónico, aparecieron quienes detectaron ciertas vulnerabilidades comerciales y tecnológicas para desarrollar diversas maneras de usufructuar con los servicios telefónicos, creando así diversas

tipologías de fraude, que han ido evolucionando y adaptándose a la par con los nuevos avances y desarrollos tecnológicos que las empresas peruanas de telecomunicaciones implementaban. La principal empresa de telecomunicaciones que presta servicios de telefonía fija es Telefónica del Perú, la cual anualmente pierde alrededor de S/. 24'000,000 por fraude.

2.1.2. FRAUDE EN LAS TELECOMUNICACIONES

Es la apropiación y/o utilización indebida e intencionada aprovechando la vulnerabilidad existente en diferentes componentes de las redes y en los procesos de sus servicios con la intención de eludir los derechos de los operadores de telecomunicaciones, para obtener beneficios económicos, ocasionando pérdidas económicas y/o de imagen a las empresas de telecomunicaciones.

Los principales motivos que se pueden dar para realizar un fraude en las telecomunicaciones son:

- Para hacer Dinero: Defraudador profesional.
- Ahorrar dinero: Defraudador casual u oportunista.
- Por fama: Hacker.
- Causar pérdida o daño: Vándalos o delincuente.

2.1.3. PRINCIPALES TIPOLOGÍAS DE FRAUDE EN TELECOMUNICACIONES

Las principales tipologías de fraude en las telecomunicaciones se asocian a todos aquellos hechos que personas ajenas a las empresas

de telecomunicaciones realizan con el fin de obtener algún beneficio, ya sea personal o económico. A continuación se detallan las modalidades de fraude externo más comunes:

2.1.3.1. Fraude por Suscripción

[SUPE01] [SCRI01] Consiste en la solicitud de nuevas líneas telefónicas y/o nuevos servicios de telecomunicaciones utilizando identidad falsa y/o utilizando la identidad de un tercero sin su consentimiento o autorización, para el uso de los servicios de telecomunicaciones y evitar cancelar la facturación generada.

Principalmente se presenta cuando se utiliza documentación falsa o de terceros con la finalidad de que los cargos y la facturación se registren a nombre de otra persona. Por lo general, cuando se adquieren servicios de telefonía, el defraudador consume lo máximo posible y posteriormente se retira del lugar donde fue instalado el servicio telefónico.

Afecta directamente a la persona suplantada quien es reportada como usuario moroso a las centrales de riesgo y a las bases de datos en el sector financiero. Por otro lado, la empresa de telecomunicaciones se ve afectada por pérdida de ingresos al no poder cobrar los consumos realizados, aumento en los reclamos, aumento de los costos, hasta llegar a ser penalizado por los entes reguladores.

La empresa debe contar con un departamento especializado en la recepción de documentos, el cual debe estar en la capacidad de detectar cuando un suscriptor está tratando de contratar servicios con documentos falsos.

La forma en que las empresas realizan la suscripción de sus clientes debe tener en cuenta los posibles fraudes que se puedan cometer. Aunque las modalidades en las que se comete el fraude cambian de acuerdo a los requerimientos que se imponen, es importante establecer parámetros que permitan evitar al máximo ser víctimas de los defraudadores.

Se pueden establecer pautas para el control de este fraude, entre las cuales tenemos:

- Contar con personal idóneo para la verificación de los documentos entregados por los clientes, capacitados en la detección e identificación de documentos falsos o adulterados.
- Establecer requerimientos mínimos para la suscripción de clientes, dependiendo del tipo de servicio que se esté solicitando. Dichos requerimientos deben tener en cuenta el tipo de cliente que se va a manejar (empresa, comercial o residencial), el valor que facturará en promedio y el tipo de tráfico que va a manejar.
- Realizar la verificación de los documentos entregados por los clientes. Dicha verificación se puede realizar ante autoridades civiles, centrales de riesgo y bases de datos propias de la empresa.

Para efectuar un análisis preventivo más detallado, tener en cuenta:

- Verificar la consistencia de los datos.
- Validar en el formato de la suscripción que todos los datos se registren de forma correcta (sin tachones y/o enmendaduras).
- Verificar el documento de identidad (DNI).

2.1.3.2. Fraude de manipulación de la central de conmutación privada (PBX)

[SUPE02] [SCRI02] Este tipo de fraude se realiza contra empresas que utilizan centrales PBX en sus instalaciones para conectarse a la red telefónica pública conmutada. Los defraudadores pueden acceder a través de los puertos DISA, la cual es una facilidad de las centrales de PBX que permite a un usuario externo realizar llamadas reoriginando desde ésta, quedando las llamadas cargadas al número del PBX. Esta opción se habilita para empleados que se encuentran fuera de la sede de la empresa y requieran usar servicios que esta tiene contratados con su proveedor, pero, debido a que estos servicios pueden ser configurados para accederse desde distintos puntos, los defraudadores que conocen el funcionamiento de la central, configuran los puertos para beneficiarse del servicio del cliente afectado.

Generalmente, las personas que realizan este fraude conocen las contraseñas para el acceso a las plantas y con ellos pueden acceder a la configuración de la PBX. Otra manera de realizar llamadas sobre los PBX, es utilizando un módem para conectarse al puerto de mantenimiento remoto. Una vez conectado, puede realizar llamadas o cambiar la configuración del PBX.

El defraudador conoce la clave de acceso para marcar sobre una extensión del PBX y realizar llamadas, ésta información se puede obtener a través de una persona de la empresa o por robo de información confidencial; para acceder al módem debe conocer el número telefónico, o la extensión que corresponde y la clave para entrar al módulo de mantenimiento del PBX.

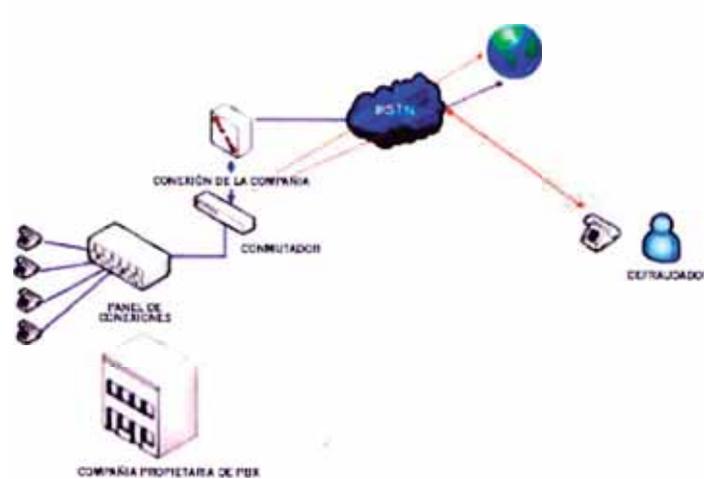


Gráfico 2.1: Estructura de la manipulación de una PBX (Fuente y Elaboración: SUPERTEL)

La detección se debe hacer a través de pruebas de la planta en su entrada a producción. Aunque la empresa en la cual se instala el PBX puede no tener el conocimiento sobre la forma en que se puede realizar el fraude, es importante contar con empresas auditoras que certifiquen este tipo de labores para evitar inconvenientes. Se debe tener información detallada de las contraseñas de configuración de la planta y el personal que tiene el conocimiento de las mismas.

Se debe verificar los destinos de las llamadas de larga distancia, y si estos corresponden a destinos que la compañía requiere para su operación. Si los consumos a estos destinos son excesivos, puede presumirse la existencia de este tipo de fraude.

La forma en la cual se programan las plantas telefónicas y los accesos troncales en las empresas es la primera medida para evitar este problema. Si se dejan circuitos abiertos para realizar

este tipo de conexiones, los costos de los servicios se van a ver afectados.

La mayoría de las empresas que tienen este tipo de tecnologías para sus servicios telefónicos, lo hacen a través de terceros, por lo tanto es necesario que se contraten empresas con respaldo y que garanticen que los servicios configurados no permitan este tipo de conexiones fraudulentas.

2.1.3.3. Hurto de Línea

[SUPE03] [SCRI03] Este tipo de fraude puede ser interno o externo, se da cuando líneas activas con asignación a usuarios son cambiadas de facilidades técnicas sin autorización del titular o de la empresa de telecomunicaciones.

Es uno de los fraudes más comunes por la facilidad de cometerlo en cualquier punto de la red externa. Generalmente esto es realizado por personal de mantenimiento, instaladores de planta externa o por ex-trabajadores de las empresas colaboradoras que conocen la distribución de la red.

Los servicios telefónicos son utilizados por terceros y el consumo es cargado al recibo mensual del titular del servicio telefónico afectado.

La acometida de la red externa al domicilio del titular es más vulnerable cuando es aérea, puede ser vulnerada desconectando el cable del domicilio y llevándolo a otro lugar, también puede realizarse en un armario al desconectar el par del abonado y empatarlo con otro que va hacia otro lugar, o entrar a una caja y romper la protección del cable troncal, derivar un cable y

conectarlo a otro del mismo cable troncal que va hacia el domicilio seleccionado por el defraudador. En sitios cerrados como edificios, se ingresan cables multipares que convergen en una caja ubicada en el sótano llamada “strip telefónico”, desde este punto se reparten las líneas de abonado mediante multipares, el infractor conecta o deriva un par telefónico en cualquiera de los puntos y/o regletas instaladas en cada uno de los pisos del edificio y usualmente cambia constantemente el abonado que está derivando para evitar ser detectado.

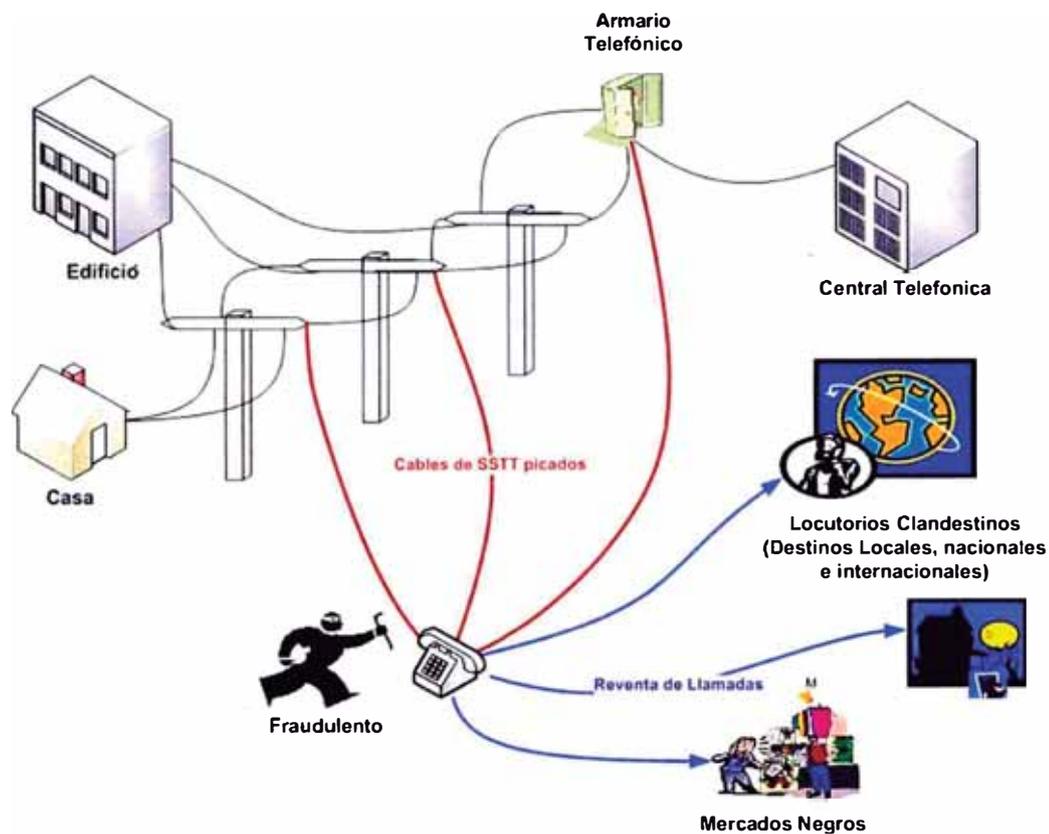


Gráfico 2.2: Estructura del hurto de línea (Fuente y Elaboración: Propia)

La empresa de telecomunicaciones se ve afectada ya que se produce un daño en la infraestructura y existe un aumento de reclamos de los clientes por averías. De igual forma este tipo de

fraude impacta directamente a la buena imagen de la compañía ante los clientes, porque el cliente se ve afectado ya que cuando no se logra comprobar un fraude y/o uso indebido del servicio telefónico por parte de un tercero, debe pagar el monto total del consumo realizado por el hurtador.

El cliente realiza un reclamo por avería al no contar con servicio, al verificarlo en los sistemas donde el servicio telefónico aparece activo; se envía una inspección técnica para realizar la revisión de la red externa, donde se encuentra la derivación o el traslado no autorizado. De igual forma, cuando un cliente que es víctima de un robo reiterado y notorio (llamadas a larga distancia internacional, larga duración, alto coste, etc.), se debe de reportar al servicio de reclamos de la empresa de telecomunicaciones para contar con el sustento ante un posible futuro reclamo del cliente.

2.1.4. EL PERFIL DEL HURTADOR DE LINEA

El Hurtador de línea es una persona con el objetivo premeditado de alcanzar un beneficio económico o propio, aprovechando las vulnerabilidades que presenta la planta externa de las empresas de telecomunicaciones, por tanto usará y combinará de manera inteligente lo siguiente:

- Facilidades técnicas expuestas.
- Concentración de múltiples servicios telefónicos en un área comercial o de alta densidad de servicios telefónicos.
- Bajos sueldos de técnicos en empresas colaboradoras (contratas y subcontratas).
- Averías e inconsistencias en los sistemas y en la planta.

- Fuga de información.
- Leyes no punibles o vacíos legales y regulatorios.

El Hurtador de línea presenta las siguientes características:

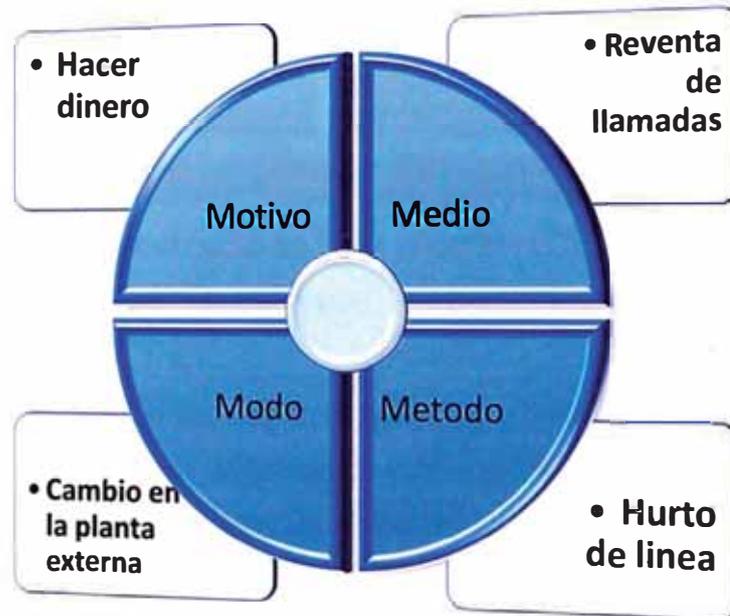


Gráfico 2.3: Ciclo de vida del fraude (Fuente y Elaboración: Propia)

2.1.4.1. Modo de Operación de un Hurtador de línea

Para poder entender la manera en que opera un hurtador de línea y poder combatirlo, se ha desarrollado la siguiente matriz donde se observa las tres preguntas básicas: ¿Qué busca?, ¿Cómo lo hace? y ¿Qué hacemos?; estas preguntas son realizadas para buscar reducir los casos de fraude por hurto de línea.



Gráfico 2.4: Matriz del hurto de línea (Fuente: LCF, Elaboración: Propia)

Como podemos observar las tres preguntas básicas se realizan sobre tres escenarios diferentes: vulnerar la planta externa, reventa de llamadas y perpetuación del uso, a continuación revisaremos cada uno.

2.1.4.1.1. Vulnerar la planta externa:



Gráfico 2.5: Ciclo de vulnerar la planta externa (Fuente: LCF, Elaboración: Propia)

El hurtador de línea busca vulnerar la planta externa en algún punto de las facilidades técnicas de un servicio telefónico, principalmente lo realizan en el armario, debido a que de realizarlo en el cable de acometida se pierde demasiada calidad en la llamada.

El hurtador de línea ya tiene identificados los servicios telefónicos dentro de un armario, cuales no cuentan con bloqueo de llamadas celulares, nacionales e internacionales, luego de esto los servicios telefónicos elegidos los conecta a puntos de prueba o los deja sin servicio, para preparar una conexión hasta el domicilio o lugar elegido para derivar los servicios telefónicos.

Las empresas de telecomunicaciones cuando un cliente indica que se encuentra sin servicio, escucha voces al levantar el auricular, no puede realizar llamadas celulares, nacionales o internacionales y no cuenta con bloqueos; envía en menos de 24 horas una inspección técnica para verificar la planta externa. En caso encuentre las facilidades técnicas alteradas las corrige y procede a realizar el seguimiento al cable de acometida del cual se derivaban las facilidades técnicas, en la mayoría de las ocasiones el otro extremo del cable se encuentra que fue cortado para no ubicar su origen, de ubicar el origen y no contar con el apoyo policial no puede realizar ninguna acción adicional.

2.1.4.1.2. Reventa de llamadas:



Gráfico 2.6: Ciclo de reventa de llamadas (Fuente: LCF, Elaboración: Propia)

El fraudulento logró derivar las facilidades técnicas de un servicio telefónico mediante la vulneración de la planta externa. Ahora el fraudulento la usa para lucrar con el fin de obtener una máxima ganancia, revende llamadas, sea directamente al público o mediante equipos que automatice el acceso. Maximiza el lucro priorizando el tráfico de larga distancia internacional (LDI) a destinos caros o tráfico a celulares (Paga quien llama – PQLL), la reventa se da en los mercados negros y lugares de alto tránsito de personas o comercios.

Los patrones de tráfico de los servicios telefónicos fraudulentos son los siguientes:

- Elevado tráfico larga distancia internacional y celular. Eventualmente larga distancia nacional (LDN).
- Generalmente los servicios telefónicos son utilizadas con fines de reventa de llamadas por lo que la dispersión de números llamados es alta.

- Si presenta tráfico LDI, se presentan llamadas a destinos con alto cargo de terminación (Cuba, Pakistán, Irak, Haití, etc.)
- Volumen de tráfico entrante es nulo comparado con el volumen de tráfico saliente.

Los controles implementados por LCF para evitar que se produzca un tráfico desmesurado de reventa y así evitar se incremente la pérdida por tráfico cursado son los siguientes:

- Monitoreo diario del tráfico PQLL, LDI y LDN de todos los servicios telefónicos.
- Reportes de consumos anómalos.
- Correlación con otros casos observados.
- Se realizan bloqueos preventivos.

2.1.4.1.3. Perpetuación del uso:



Gráfico 2.7: Ciclo de perpetuación de uso (Fuente: LCF, Elaboración: Propia)

La perpetuación del uso la realiza debido a que el hurtador de línea ya tiene mapeado los armarios o puntos de planta externa más vulnerables en determinada zona y se va moviendo dentro de esta zona hurtando la línea a diferentes servicios telefónicos en diferentes puntos para no ser detectado. Existen mafias en zonas de alto índice delincriminal que no permiten a los técnicos reparar las averías bajo amenazas de atentar contra su integridad.

Para evitar que las pérdidas sean cuantiosas al perpetuar un servicio telefónico, las empresas de telecomunicaciones en zonas donde el índice de hurto de línea es muy elevado, proceden a reforzar la protección en los armarios instalando dobles abrazaderas con candados reforzados, jaulas de protección para el armario, etc., sin embargo es un costo adicional en el que tendría que incurrir la empresa de telecomunicaciones.

2.2. JEFATURA DE LUCHA CONTRA EL FRAUDE SUPERVISION RED FIJA

La Jefatura de Lucha contra el fraude (LCF) pertenece a la Dirección de Seguridad y Protección de **tggestiona**; dentro de LCF se encuentra la Supervisión de Red Fija que es el área especializada que se encarga del análisis de los diversos casos de fraude que se generan en los servicios de cobre.

LCF le brinda a la principal empresa de telecomunicaciones peruana el Servicio de Control de Fraude en sus diferentes modalidades.

2.2.1. SERVICIO DE CONTROL DE FRAUDE

El Servicio de Control de Fraude consiste en la protección de ingresos de la principal empresa de telecomunicaciones peruana, ante la eventualidad de fraude en telecomunicaciones perpetrado por personal interno, malos clientes, entidades y/o personas fraudulentas en los diferentes servicios que se ofrecen, minimizando el impacto en los aspectos Económico y Financieros, mediante la prevención, detección, análisis, solución y disuasión para enfrentar las diferentes tipologías de Fraude en Telecomunicaciones.

En resumen el servicio de Control de fraude que brinda LCF se puede resumir en la siguiente frase: “Evitamos pérdida económica y de imagen a nuestros clientes”.

En el siguiente gráfico se muestran las diferentes fases que se desarrollan para la lucha contra el fraude y que acciones se toman en cada uno de ellas.



Gráfico 2.8: Fases de la lucha contra el fraude (Fuente: LCF, Elaboración: Propia)

De las diversas fases de la lucha contra el fraude se listan a continuación algunas de las acciones que se desarrollan:

- Identificación de nuevas modalidades de fraude.
- Elaboración de recomendaciones por modalidad a las diferentes unidades organizativas.
- Elaboración de medidas preventivas ante los riesgos de fraude.
- Elaboración de Estadísticas de incidencias en los distintos tipos de fraude.
- Monitoreo y análisis de alertas de tráfico en la red fija.
- Determinación de clientes y servicios de prestación de alto riesgo por fraude en el tiempo de detección.

- Establecimiento de procesos permanentes con las respectivas áreas de la empresa cliente.
- Determinación de parámetros de umbrales sobre las cuales se determinaran los casos.
- Proveer información requerida para operativos de LCF.
- Determinación de estrategias a seguir.
- Presentación de informes mensuales de los casos gestionados por fraude.

En el gráfico mostrado se observa como interactúa el Servicio de control de fraude con las diversas áreas de la principal empresa de telecomunicaciones peruana, tenemos a:

- Back Office Comercial que se encarga de la revisión de las altas nuevas generado por los diversos canales de venta.
- Instalaciones que gestiona toda la planta técnica, está encargado de instalar las altas nuevas de los servicios telefónicos.
- TUP está encargado de la gestión de los teléfonos de uso público, TPE: Teléfono Publico Externo y TPI: Teléfono Publico Interno.
- Back Office Reclamos se encarga de la revisión de los expedientes de los diversos reclamos presentados.
- Regulación se encarga de la parte legal evitando incurrir en alguna falta o infracción que incurran en una multa o sanción a la empresa.
- Facturación y Cobranzas se encarga de la emisión de facturas y el posterior seguimiento para la cobranza de la misma.
- GORAC se encarga de realizar las inspecciones técnicas en caso de avería, reporte de reclamos de facturación elevado desconociendo tráfico o a solicitud de LCF.

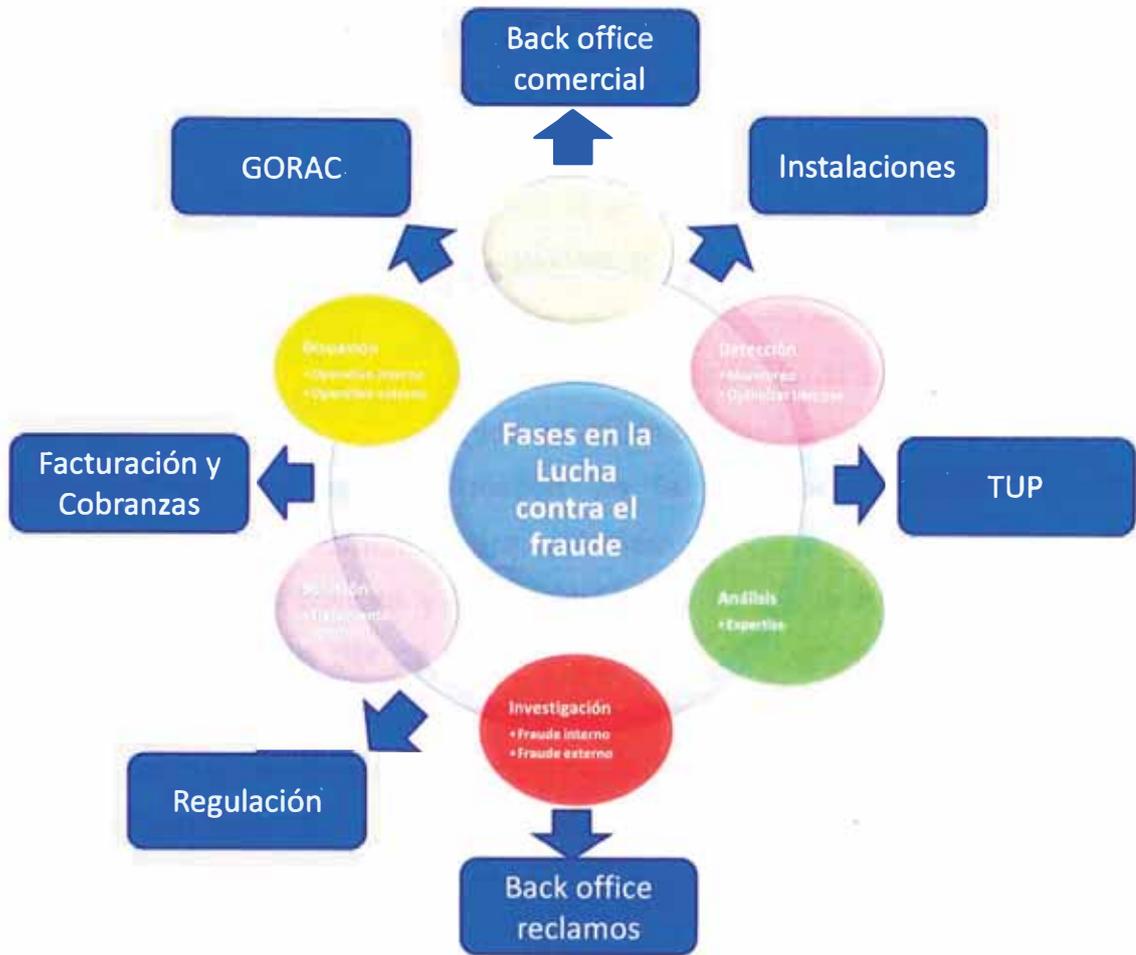


Gráfico 2.9: Interrelación del servicio de control de fraude (Fuente: LCF, Elaboración: Propia)

El Servicio de control de fraude presenta algunas debilidades que no le permiten trabajar con la efectividad deseada y que limitan su accionar, entre ellos podemos mencionar los siguientes puntos:

- Falta de normas regulatorias y legales específicas para calificar apropiadamente determinadas modalidades de fraude en las telecomunicaciones; de tal forma que permita neutralizar y/o sancionar penalmente este tipo de actividades delictivas.
- Falta de funcionalidades en los sistemas comerciales que permitan una mejor identificación de los casos observados por fraude así como la implementación de reglas más preventivas.

- Falta de sinergias entre áreas de la empresa cliente y su soporte que permitan un combate más eficiente del fraude.

2.3. MINERIA DE DATOS

La definición del concepto de Minería de Datos puede variar entre unos investigadores y otros. Por ejemplo, los estadísticos, analistas de datos y la comunidad de sistemas de gestión de la información adoptan mayoritariamente este término para referirse al proceso genérico correspondiente a las técnicas y herramientas de investigación usadas para extraer información útil de una base de datos. Dentro de estas técnicas podemos considerar todos aquellos métodos matemáticos y técnicas software para el análisis inteligente de los datos y búsqueda de patrones o tendencias en los mismos aplicados de forma iterativa e interactiva.

El nombre de Minería de Datos, deriva de la similitud que se encuentra entre buscar valiosa información de negocios en grandes bases de datos con la búsqueda de vetas de metales preciosos dentro de una montaña. Ambos procesos requieren examinar inteligentemente una inmensidad de material hasta encontrar algo que pueda resultarnos útil y valiosa.

Dentro de las definiciones que se pueden encontrar en la literatura relacionada se muestran algunas de las más significativas:

- “Minería de Datos es la exploración y análisis, mediante métodos automáticos o semiautomáticos, de grandes cantidades de datos para descubrir reglas o patrones significativos”.
- “Minería de Datos es el proceso analítico diseñado para explorar grandes cantidades de información con el fin de investigar patrones

consistentes y/o relaciones sistemáticas entre variables y, a continuación, validar los resultados aplicando modelos detectados para nuevos subgrupos de datos”.

- “Minería de Datos es el conjunto de técnicas y herramientas aplicadas al proceso trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con el objeto de predecir de forma automatizada tendencias y comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos”.

A grandes rasgos se puede decir que la Minería de Datos es un conjunto de técnicas y herramientas que permiten extraer el conocimiento útil (patrones de comportamiento, modos de operación, información útil para descubrir fallos, tendencias, etc.) para la ayuda en la toma de decisiones, comprensión y mejora de procesos o sistemas, etc.; partiendo de grandes cantidades de datos.

Para alcanzar buenos resultados es necesario comprender que la Minería de Datos no se basa en una técnica estándar y genérica que resuelve todo tipo de problemas, sino que consiste en una técnica dinámica e iterativa que va a depender del problema planteado, de la disponibilidad de la fuentes de datos, del conocimiento de las herramientas necesarias y de los requerimientos y recursos de la empresa.

2.3.1. FASE DE UN PROCESO CLASICO DE MINERIA DE DATOS

Las fases en el proceso global de Minería de Datos no están claramente diferenciadas lo que hace que sea un proceso iterativo e interactivo con el usuario experto. Las interacciones entre las

decisiones tomadas en diferentes fases, así como los parámetros de los métodos utilizados y la forma de representar el problema suelen ser extremadamente complejos. Pequeños cambios en una parte pueden afectar fuertemente al resultado final. Se puede considerar la estructura del proceso en seis pasos definidos en el Gráfico siguiente:

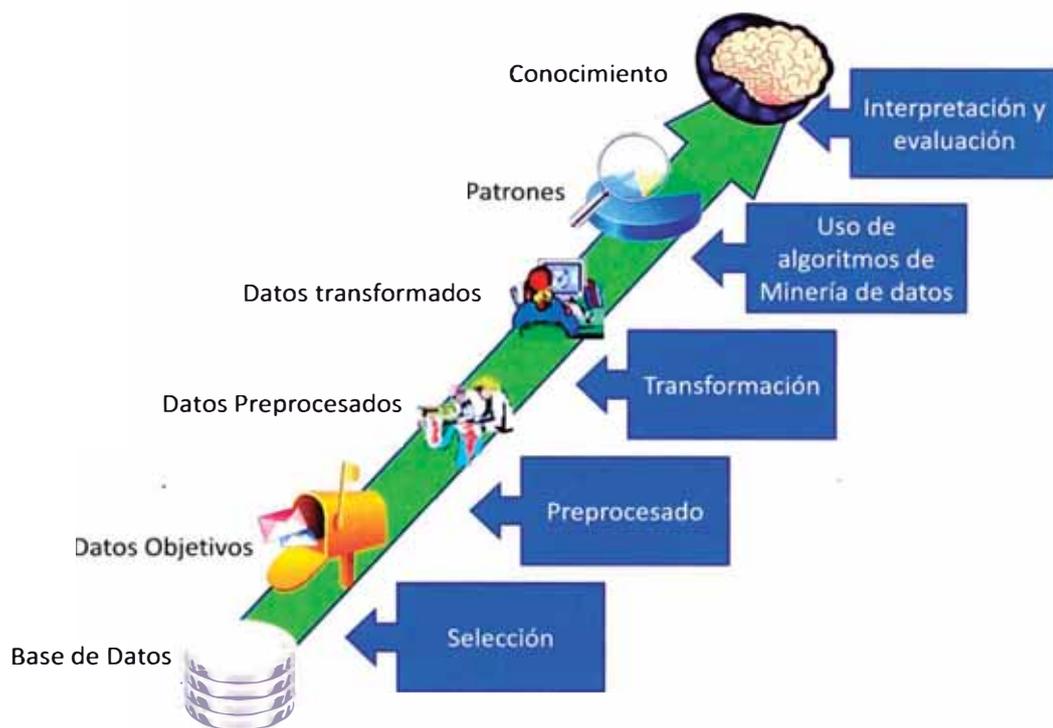


Gráfico 2.9: Fases típicas de un proceso de Minería de Datos y KDD
(Fuente Kmining.com, Elaboración: Propia)

2.3.1.1. Definición del Alcance y Objetivos

El primer paso de un proyecto de Minería de Datos consiste en conocer el desarrollo y dominio de la aplicación, determinar el conocimiento relevante a usar, así como establecer los objetivos del usuario final. El desarrollo de esta primera fase, establece las bases para la realización de las posteriores fases del proceso. Por lo que el éxito o fracaso del proceso va a depender en gran medida de las decisiones que se adopten en esta etapa. En esta

fase se determinan los factores que son susceptibles de un procesado automático, los cuellos de botella del dominio, los conocimientos a priori que se tienen del proceso, así como cuáles son los objetivos finales que se pretenden lograr y cuáles van a ser los criterios de rendimiento exigibles. Por lo tanto, esta fase requiere cierta dependencia usuario-analista, siendo necesario el establecimiento de unos canales de comunicación entre ambas partes.

Generalmente, es conveniente volcar todos los esfuerzos iniciales en comprender el sistema en todos sus detalles, para evitar llegar a callejones sin salida debido a una falta de comprensión de algunas de las partes del proceso.

[DOPY01] Dorian Pyle considera que un 80% de la importancia para llegar al éxito proviene en la forma de abordar el problema, definir cuáles pueden ser las pautas para llegar a la solución y la forma de implementarlas para solucionar el problema con éxito.

Tarea	Tiempo dedicado (%)	Importancia para llegar al éxito (%)
1. Definir el problema	10%	15%
2. Explorar la solución	9%	14%
3. Implementación de los resultados	1%	51%
4. Preparación de los datos	60%	15%
5. Procesamiento de los datos	15%	3%
6. Modelado y testeo de los datos	5%	2%

Cuadro 2.1: Tiempo e importancia de cada una de las fases

(Fuente: Dorian Pyle, Elaboración: Propia)

Fundamentalmente, según Pyle, los pasos que llevan al éxito van a depender de los siguientes aspectos:

2.3.1.1.1. *Identificación correcta de los problemas a resolver.*

Muchas veces, esta tarea puede parecer trivial, pero puede suceder que el problema no se comprenda completamente. Existe el caso de una empresa de telecomunicaciones que pretendía mejorar un modelo de predicción de los clientes que tenían un alto porcentaje de probabilidades en darse de baja. Después de pasar por alto esta primera fase, ya que consideraban que tenían muy bien definido el problema, se desarrolló un modelo predictivo que parecía tener una eficiencia del 80% frente al 50% del modelo anterior y se planteó una campaña de marketing, con un elevado gasto de dinero, dirigida hacia esos clientes que, según el modelo, podían darse de baja en poco tiempo. El resultado final de esa campaña de marketing fue desastroso, ya que, por ejemplo, se habían considerado dentro del modelo que, personas desempleadas mayores de 80 años tenían muchas posibilidades en anular su contrato con la empresa. Y, claro está, muchas de estas personas morían y por lo tanto, ningún programa de incentivos ni de buzoneo podía evitar que estas personas se dieran de baja de la compañía de telecomunicaciones. Esto, condujo a pensar a la empresa, que el problema había estado mal definido desde el principio y que había que considerarlo según diferentes segmentos de mercado.

2.3.1.1.2. *Definición con precisión de los problemas.*

Será necesario dividir las descripciones del problema que son demasiado generalistas en componentes más pequeños que puedan ser contrastados por la información examinada. Por

ejemplo, un problema del tipo: “Me gustaría tener un modelo matemático de los índices de fallos que se producen en mi línea de ensamblaje los lunes y los viernes para poder eliminarlos o reducirlos...”, puede ser necesario recomponerlo en pequeños y más abordables subproblemas del tipo: “detectar los tipos de fallos que se producen”, “por qué los lunes y viernes sobresalen de los demás días”, “determinar qué hay que examinar (empleados, maquinaria, etc.)”; “cuáles son las piezas donde se producen los fallos”, etc.

2.3.1.1.3. Resolver las ambigüedades.

Es conveniente resolver las ambigüedades que puedan surgir debido a que la imagen mental del problema en la propia mente del cliente está formada por una gran cantidad de conceptos asociados, que él tiene asumidos, pero que probablemente, pueden no ser tan claros para aquellas personas que no conocen el proceso con profundidad, además se debe determinar, dentro del número de problemas, el grado de importancia y dificultad de cada uno de ellos.

2.3.1.1.4. Definir qué resultados se esperan conseguir.

También resulta de especial importancia determinar el tipo de resultado que se busca: un modelo matemático, unas gráficas, unos informes, etc.; definiéndolo de la forma más completa posible. Esto permite que las tareas posteriores se dirijan directamente hacia el objetivo buscado.

2.3.1.1.5. Implementar los resultados obtenidos.

Para conseguir el éxito, no solo es importante obtener unos resultados sino también tener en cuenta la forma con la que se aplicarán.

Como es lógico, toda esta información será desarrollada y contrastada con diferentes expertos hasta alcanzar un consenso entre las partes implicadas, pudiéndose utilizar algunas de las conocidas técnicas de consultas a expertos.

2.3.1.2. Selección de los Datos Relevantes

La identificación de los datos relevantes para una operación de Minería de Datos es una tarea que no puede ser automatizada y que por lo tanto debe ser realizada por el analista. Esta tarea consiste en la creación del conjunto de datos objetivo, enfocando la búsqueda en subconjuntos de variables y/o muestras de datos en donde realizar el proceso de análisis.

En esta fase deben de ser seleccionados, de forma coordinada por el analista y el usuario, los datos más relevantes del proceso, así como su disponibilidad. Esto implica consideraciones sobre la homogeneidad y variación a lo largo del tiempo de los datos, los grados de libertad o la estrategia de muestreo.

La obtención de datos puede realizarse directamente desde sistemas transaccionales, archivos o a partir de un almacén de datos (Data Warehouse). Habitualmente esta obtención viene predeterminada en función de la disponibilidad de los datos: existencia de bases de datos para el proceso, datos almacenados en archivos, necesidad de implantar un nuevo sistema de adquisición de datos, etc. El tomar datos a partir de sistemas de

operación puede suponer el riesgo de encontrarse con problemas de ruido o sin datos.

2.3.1.3. Preprocesado y Limpieza de Datos

El objetivo del preprocesado de datos es la transformación del conjunto original de datos en un nuevo conjunto de datos más significativo y manejable. Un 60% del tiempo se dedica al preprocesado de los datos.

El preproceso es una transformación T que transforma la matriz que contiene los datos reales del proceso, X , en una nueva matriz Y tal que:

- Y conserva la información de X .
- Y elimina al menos uno de los problemas contenidos en X .
- Y es más útil que X

El preproceso de los datos incluye cuatro etapas principales: Identificación y conversión de tipos, imputación (rellenar los datos inexistentes), identificación de espurios (outliers), eliminación de ruido y datos incompletos.

2.3.1.3.1. Identificación y conversión de atributos

Las primeras tareas de preprocesado, son las más arduas ya que, generalmente, deben consistir en identificar, casi manualmente, los diferentes tipos de atributos existentes en la base de datos y convertirlos a otro tipo dependiendo de las necesidades posteriores. Fundamentalmente, podemos clasificarlos en los siguientes dos grupos:

- Numéricos o Cuantitativos. También algunas veces llamados “continuos”.
- Nominales o Cualitativos. También algunas veces llamados “discretos”. Aunque la literatura estadística introduce unos “niveles de medida” clasificados en los siguientes subgrupos:
 - Nominales. Que corresponden a valores que tienen distintos símbolos generalmente denominados etiquetas o nombres. Por ejemplo: colores. Un caso especial son los datos binarios (que solo pueden tener dos valores).
 - Ordinales. Que determinan un cierto ranking en las categorías. Por ejemplo: frío < templado < caliente, bajo < medio < alto, etc.
 - Intervalos. Que son valores que no solo están ordenados sino también medidos en unidades iguales con un cero arbitrario. Por ejemplo: temperaturas, años, etc.
 - Ratios. Que corresponden con medidas donde está definido un punto cero inherente en sí mismo. Por ejemplo: la distancia de un objeto a otro, tiene como cero la distancia del objeto a sí mismo, temperatura en grados absolutos, edad desde el Big Bang, etc.

Los atributos, según el tipo que sean, deben ser acomodados a los algoritmos que se vayan a utilizar. De esta forma, muchas veces resulta necesaria la conversión de los datos para que puedan ser tratados convenientemente. Este proceso de conversión depende en gran manera de los esquemas utilizados. Por ejemplo, algunos de los esquemas utilizan valores formados por escalas ordinales y solamente usan comparaciones mayor-que, menor-que para compararlos. Otros

en cambio, usan escalas tipo ratios y usan distancias entre ellas. Es decir, es necesario comprender cómo trabajan los algoritmos de minería de datos que vamos a utilizar para saber cómo preparar los datos.

2.3.1.3.2. Conversión de Tipos de Variables

Dependiendo de los algoritmos que vayamos a utilizar deberemos transformar los datos de un tipo a otro. Por ejemplo, un atributo nominal no puede ser tratado por una red neuronal o un clasificador basado en árboles puede necesitar que los datos sean nominales.

Muchas veces un atributo nominal puede ser convertido a un atributo ordinal simplemente indicándole al sistema unas reglas que relacionen éstos. Por ejemplo, una serie de la forma: {bajo, alto, medio}, fácilmente puede ser convertida en una serie ordinal simplemente mediante la regla: bajo < medio < alto, o alto > medio > bajo, aunque otras veces las reglas no son tan claras.

La conversión de datos nominales a numéricos dependerá del conocimiento que tengamos sobre el grado de cercanía o alejamiento de unos con otros. Por ejemplo, si tenemos una serie de datos del tipo: {error insignificante, error medio, error peligroso}, y queremos alimentar con ellos una variable numérica de un modelo matemático, será necesario desarrollar una escala de medidas numéricas que se adapten convenientemente.

Un caso más interesante, es la conversión de una serie de valores numéricos a una serie de datos nominales. Esta

transformación consistirá fundamentalmente en la creación de clases agrupando los conjuntos de datos según algún criterio preestablecido: distancia, similitud, en relación a otra variable, etc.

Otro tipo de transformaciones más avanzadas se basan en reglas fuzzy o difusas, capaces de tratar las incertidumbres mediante funciones aplicadas a cada valor del campo.

Una vez tenemos los tipos de atributos adaptados a nuestras necesidades, será conveniente realizar las siguientes fases:

- Detectar los espurios y eliminarlos.
- Rellenar los datos inexistentes.
- Eliminar el ruido.

2.3.1.4. Transformación de los Datos

La fase de transformación y reducción de los datos, es otra de las fases críticas dentro del proceso global que necesita de un buen conocimiento y una buena intuición que determinará el éxito o el fracaso del proceso de Minería de Datos.

Se busca, por un lado, preparar la información que se tiene para que pueda ser procesada por los algoritmos de minería de datos y además, reducir la cantidad de información redundante para simplificar las tareas posteriores. Se busca por lo tanto:

- Extracción de las características o atributos útiles de los datos (reducción de dimensionalidad).

- Transformación de los datos con el objetivo de proporcionar una representación de los datos más intuitiva y manejable.
- Fundamentalmente podemos destacar tres tareas específicas:
 - Reducción de los Datos.
 - Creación de Datos Derivados.
 - Transformación de la distribución de los Datos.

2.3.1.5. Uso de los Algoritmos de Minería de Datos

Una vez obtenidos los datos transformados y preparados en una base de datos normalizada, con variables poco correlacionadas entre sí, con los espurios, el ruido eliminado y con una dimensión adecuada; sería el momento del uso de los algoritmos de minería de datos.

Las herramientas de Minería de Datos empleadas en el proceso de extracción de conocimiento se pueden clasificar en dos grandes grupos:

- Técnicas de verificación (en las que el sistema se limita a comprobar hipótesis suministradas por el usuario).
- Métodos de descubrimiento (en los que se han de encontrar patrones potencialmente interesantes de forma automática, incluyendo en este grupo todas las técnicas de predicción). El resultado obtenido con la aplicación de algoritmos de descubrimiento puede ser de carácter descriptivo o predictivo. Las predicciones sirven para prever el comportamiento futuro de algún tipo de entidad mientras que una descripción puede ayudar a su comprensión.

Antes de poder utilizar los datos, casi siempre es necesario preprocesarlos para adecuarlos a las necesidades de las técnicas que se van a utilizar sobre ellos. Las técnicas de visualización son muy útiles en este momento, para aumentar el conocimiento previo de los datos y como paso previo a procesos posteriores. También ayudan a descubrir la estructura de clúster de los datos y posibles correlaciones entre ellos; así como en la detección de espurios.

Los algoritmos de minería de datos pueden ser utilizados para alguna de las siguientes tareas:

- **Agrupamiento o segmentación:** Se busca la identificación de tipologías o grupos en los cuales los elementos guardan similitud entre sí y se diferencian de los otros grupos. Esto permite el tratamiento particularizado de cada una de estas agrupaciones.
- **Asociación:** Consiste en establecer las posibles relaciones entre acciones o sucesos aparentemente independientes. Así, se puede reconocer cómo la ocurrencia de un determinado suceso puede inducir la aparición de otro u otros.
- **Secuencia:** Es un concepto similar al anterior, pero en el que se incluye el factor tiempo. Es decir, permite reconocer el tiempo que transcurre o suele transcurrir entre el suceso inductor y los sucesos inducidos.
- **Reconocimiento de patrones:** Se trata de analizar la asociación de una señal o información de entrada con aquella o aquellas con las que guarda mayor similitud, y que están ya catalogadas en el sistema. Generalmente se usan para identificar las causas de problemas o incidencias y buscar las posibles

soluciones, siempre y cuando se disponga de la base de información necesaria en la que buscar.

- **Previsión:** Se busca establecer el comportamiento futuro más probable de una variable o una serie de variables a partir de la evolución pasada y presente de las mismas o de otras de las cuales dependan. Las técnicas asociadas a estas herramientas tienen ya un elevado grado de madurez.
- **Simulación:** Comparan la situación actual de una variable y su posible evolución futura según la variación probable de las que depende.
- **Optimización:** resuelve el problema de la minimización o maximización de una función que depende de una serie de variables, encontrando los valores de éstas que satisfacen la condición de máximo (típicamente beneficios), o mínimo (típicamente costos). Normalmente suele haber unas restricciones, que hacen que no todas las posibles soluciones sean aceptables, de modo que el universo de búsqueda se reduce a aquellas soluciones que satisfagan las restricciones.
- **Clasificación:** Agrupa a todas las herramientas que permiten asignar a un elemento la pertenencia a un determinado grupo o clase. Esto se lleva a cabo a través de la dependencia de la pertenencia a cada clase en los valores de una serie de atributos o variables. Se establece un perfil característico de cada clase y su expresión, en términos de un algoritmo o reglas, en función de las distintas variables. Se establece también el grado de discriminación o influencia de estas últimas. Con ello, es posible clasificar un nuevo elemento una vez conocidos los valores de las variables presentes en él.

Para desarrollar todos estos procesos, se dispone de una extensa gama de técnicas que le pueden ayudar en cada una de las fases de dicho proceso.

2.3.1.6. Interpretación de los Resultados

La interpretación y verificación de resultados es un proceso complejo. La obtención de resultados aceptables dependerá de factores como: definición de medidas del interés del conocimiento (de tipo estadístico, en función de su sencillez) que permitan filtrarlo de forma automática, existencia de técnicas de visualización para facilitar la valoración de los resultados o búsqueda manual de conocimiento útil entre los resultados obtenidos.

Un factor muy importante en esta fase, es el grado de experiencia y conocimiento del analista. La cantidad de información extraída depende en gran medida del grado de conocimiento que el analista tenga del problema, así como de sus experiencias en la resolución de problemas similares.

Las decisiones tomadas durante esta fase irán encaminadas en dos direcciones:

2.3.1.6.1. Verificación de resultados:

La verificación de resultados incluye determinar el grado de cumplimiento de los objetivos finales establecidos durante la primera fase del proceso de Minería de Datos, así como la validación de la información extraída. Durante esta fase se debe verificar la coherencia de la información obtenida con otros tipos de conocimiento ya previamente asentado y

aceptado, resolviendo las posibles inconsistencias existentes. Si los objetivos finales han sido alcanzados, se procederá a la consolidación del conocimiento descubierto, incorporándolo al sistema o simplemente documentándolo y enviándolo a la parte interesada. En caso contrario se procederá a la obtención de más información.

2.3.1.6.2. Obtención de más información:

La información extraída se utilizará como información a priori para la extracción de más información. Para ello será necesario retornar a alguna de las fases anteriores del proceso de Minería de Datos y modificar algunas de las decisiones tomadas durante esas fases, haciendo para ello uso de la nueva información obtenida. De esta forma el proceso de Minería de Datos se convierte en un proceso potencialmente iterativo. Algunas de las decisiones que pueden ser tomadas para la obtención de más información son, por ejemplo: recolección de nuevos datos, separación de datos en clases, transformaciones de las variables, eliminación de datos, selección de otros algoritmos de Minería de Datos; cambio en los parámetros introducidos en los algoritmos, delimitación del campo de búsqueda, etc.

2.3.2. METODOLOGÍAS DE APLICACIÓN DE LA MINERÍA DE DATOS

Las metodologías más aplicadas a nivel mundial son CRISP-DM y SEMMA. Las razones fundamentales son debidas a su generalización y practicidad, además de su libre utilización [KDNU01].

Metodología	%
CRISP-DM	42%
Metodología Propia	19%
SEMMA	13%
Proceso KDD	7%
Organización Propia	5%
Metodología de Dominio Especifico	5%
Otras metodologías, de dominio no específico	4%
Ninguna	5%

Cuadro 2.2: Encuesta sobre la metodología utilizada (Fuente: kdnuggets.com, Elaboración: Propia)

Un punto en común que se puede observar en todas estas metodologías es que todas se basan en un modelo espiral, de forma que se retorna a las primeras fases del proceso pero a un nivel superior ya que la comprensión alcanzada es mayor. Además, en todas vemos la importancia del análisis inicial del contexto y de la validación final de la toma de decisiones.

A la vista de las dificultades anteriores y para utilizar estas técnicas de forma eficiente y ordenada, es preciso aplicar una metodología estructurada. A este respecto se proponen las siguientes metodologías, siempre adaptables a la situación a la que se aplique.

2.3.2.1. Metodología CRISP-DM

[SPSS01] CRISP-DM (CROSS-Industry Standard Process for Data Mining), es una metodología para el desarrollo de proyectos de minería de datos que se ha convertido en un estándar de facto. El consorcio CRISP-DM, responsable de esta metodología, está integrado por importantes empresas europeas y estadounidenses que poseen una amplia experiencia en proyectos de análisis de datos relacionados con muy diversos campos de la industria.

La metodología para minería de datos CRISP-DM, está descrita como un proceso jerárquico, que consiste en un conjunto de tareas descritas en cuatro niveles de abstracción, desde el general hasta el específico: fase, tareas generales, tareas específicas e instancias de proceso.

El proceso está organizado en un número de fases; cada fase consiste en varias tareas generales de segundo nivel. Este segundo nivel se denomina genérico porque se pretende que sea lo suficientemente general como para cubrir todas las posibles situaciones. Las tareas generales deben ser lo más completas y estables posibles. Se entenderán tareas completas a aquellas que cubran completamente el proceso de análisis y sus posibles aplicaciones. Por otro lado, se entiende como estables aquellas tareas que cubran incluso desarrollos aún no conocidos. El tercer nivel, el nivel de tareas especializadas, es el lugar en el que se describe cómo las acciones de las tareas generales (nivel 2) se deberían desarrollar en ciertas situaciones específicas. Por ejemplo, en el segundo nivel puede existir una tarea general llamada "limpieza de datos". El tercer nivel describe cómo difiere esta tarea de unas situaciones a otras, por ejemplo, la limpieza de valores numéricos y la limpieza de valores categóricos o si el tipo de problema es un clusterizado a un modelo predictivo.

La descripción de fases y tareas en pasos discretos desarrollados en un orden específico representa una secuencia idealizada de eventos. En la práctica, muchas de estas tareas pueden ser desarrolladas en un orden diferente y frecuentemente será necesario volver atrás a tareas previas y repetir ciertas acciones. El modelo de procedimiento no pretende abarcar todas estas

posibles rutas a lo largo del proyecto porque esto requeriría un modelo enormemente complejo.

El cuarto nivel, el nivel de instancias de proceso, es un conjunto de acciones, decisiones y resultados sobre el proceso de minería de datos en curso. Una instancia de proceso se organiza de acuerdo con las tareas definidas en los niveles superiores, pero representa lo que pasa en realidad en un proceso particular, más que lo que pasa en general.

Horizontalmente, la metodología CRISP-DM distingue entre el modelo de referencia y la guía del usuario. El modelo de referencia presenta una vista rápida de las fases, tareas y sus salidas y describe lo que hay que hacer en un proyecto de minería de datos. La guía del usuario da consejos y trucos mucho más detallados para cada fase y para cada tarea dentro de una fase y describe cómo desarrollar un proyecto de análisis de datos.

CRISP-DM define las diferentes fases de las que consta un proyecto, las tareas correspondientes y las relaciones entre ellas.

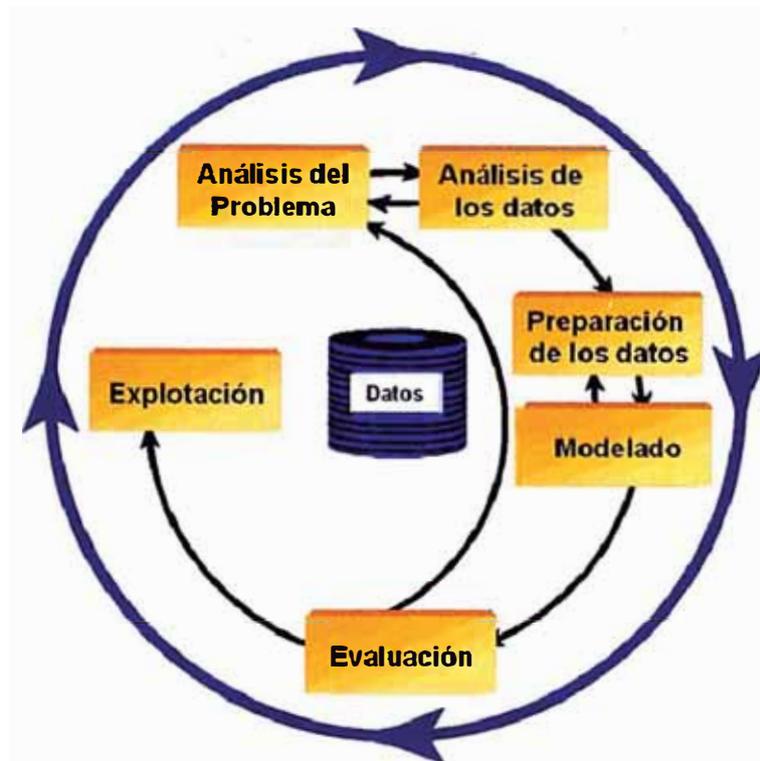


Gráfico 2.10: Fases de la metodología CRISP-DM (Fuente y Elaboración: monografias.com)

En la Figura se muestran las 6 fases definidas. El orden de las mismas no es estricto, ya que frecuentemente a lo largo del desarrollo del proyecto, es necesario volver atrás en numerosas ocasiones, dependiendo de los resultados obtenidos en las fases previas. Las flechas indican las relaciones más habituales entre las fases. El círculo exterior simboliza la naturaleza cíclica de la Minería de Datos, ya que la solución a la que finalmente se llega puede conducir al planteamiento de nuevas cuestiones que den origen a otros proyectos. A continuación, se resumen las tareas genéricas en las que se desglosan cada una de las fases y las salidas generadas por cada una de ellas.

- **Análisis del Problema:** Fase inicial que incluye la comprensión de los objetivos y requerimientos del proyecto

desde una perspectiva de negocio, con el fin de convertirlos en objetivos y en una planificación.

- **Análisis de los Datos:** Recolección inicial de datos para familiarizarse con ellos, identificar su calidad y descubrir las relaciones entre los más evidentes para las primeras hipótesis de relaciones ocultas entre ellos.
- **Preparación de los Datos:** Construcción de la base de datos a partir de los datos primarios. Se desarrolla repetitivamente y no de una forma muy estructurada. Incluye la selección de tablas, registros y atributos, además la transformación y preparación para las herramientas de modelado.
- **Modelado:** Se seleccionan y aplican varias técnicas. Normalmente existen varias técnicas para el mismo problema y cada una exige una entrada de datos particular por ello es necesario interactuar con la fase anterior para adecuar la base de datos de trabajo. Los parámetros son calibrados.
- **Evaluación:** Una vez creado un buen modelo se debe evaluar el rendimiento del mismo y la integridad de todos los pasos sobre todo teniendo en cuenta que se han introducido todos los criterios de negocio. Se debe dar el visto bueno final a la aplicación del modelo de minería de datos.
- **Desarrollo:** Normalmente los proyectos de minería de datos no terminan en la implantación del modelo sino en el incremento de conocimiento obtenido de los datos. Para ello es imprescindible documentar y presentar los resultados de manera comprensible.

Además debe asegurarse el mantenimiento de la aplicación y la posible difusión de estos resultados.

2.3.2.2. Metodología SEMMA

[SASS01] SAS Institute desarrollador de esta metodología, la define como el proceso de selección, exploración y modelado de grandes cantidades de datos para descubrir patrones de negocio desconocidos. El nombre de esta terminología es el acrónimo correspondiente a los cinco pasos básicos del proceso “Sample, Explore, Modify, Model and Assess”. El esquema siguiente presenta la dinámica del sistema.

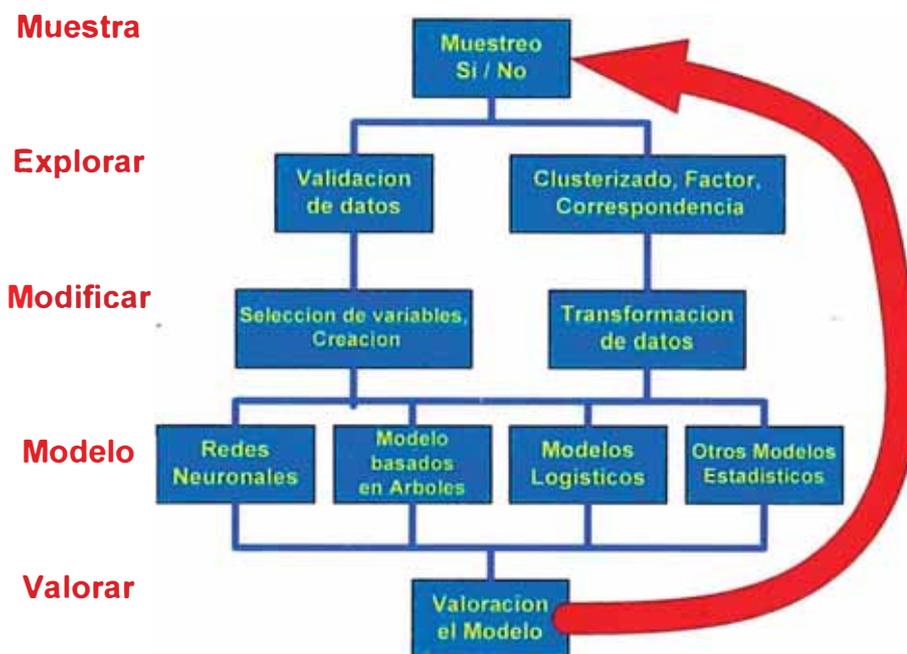


Gráfico 2.11: Fases de la metodología SEMMA (Fuente: dataprix.com, Elaboración: Propia)

2.3.2.2.1. Muestreo

Extracción de la población muestral sobre la que se va a aplicar el análisis. En ocasiones se trata de una muestra aleatoria, pero puede ser también un subconjunto de datos del repositorio de datos que cumplan unas condiciones determinadas. El

objeto de trabajar con una muestra de la población en lugar de con toda ella, es la simplificación del estudio y la disminución de la carga de proceso.

La muestra más óptima será aquella que, teniendo un error asumible, contenga el número mínimo de observaciones.

En el caso de que se recurra a un muestreo aleatorio, se debería tener la opción de elegir entre:

- El nivel de confianza de la muestra (usualmente el 95% o el 99%).
- El tamaño máximo de la muestra (número máximo de registros), en cuyo caso el sistema deberá informar del error cometido y la representatividad de la muestra sobre la población original.
- El error muestral que está dispuesto a cometer, en cuyo caso el sistema informará del número de observaciones que debe contener la muestra y su representatividad sobre la población original.

Para facilitar este paso se debe disponer de herramientas de extracción dinámica de información con o sin muestreo (simple o estratificado). En el caso del muestreo, dichas herramientas deben tener la opción de, dado un nivel de confianza, fijar el tamaño de la muestra y obtener el error o bien fijar el error y obtener el tamaño mínimo de la muestra que proporcione este grado de error.

2.3.2.2.2. Exploración

Una vez determinada la población que sirve para la obtención del modelo se deberá determinar cuáles son las variables explicativas que van a servir como entradas al modelo. Para ello es importante hacer una exploración de la información disponible de la población que permita eliminar variables que no influyen y agrupar aquellas que presentan efectos similares.

El objetivo es simplificar en lo posible el problema con el fin de optimizar la eficiencia del modelo. En este paso se pueden emplear herramientas que permitan visualizar de forma gráfica la información, utilizando las variables explicativas como dimensiones.

También se pueden emplear técnicas estadísticas que ayuden a poner de manifiesto relaciones entre variables. A este respecto resultará imprescindible una herramienta que permita la visualización y el análisis estadístico integrado.

2.3.2.2.3. Manipulación

Tratamiento realizado sobre los datos de forma previa a la modelado, en base a la exploración realizada, de forma que se definan claramente las entradas del modelo a realizar (selección de variables explicativas, agrupación de variables similares, etc.).

2.3.2.2.4. Modelado

Permite establecer una relación entre las variables explicativas y las variables objeto del estudio, que posibilitan inferir el valor de las mismas con un nivel de confianza determinado.

2.3.3. APLICACIÓN DE LA MINERÍA DE DATOS Y TENDENCIAS

Hasta ahora esta técnica ha sido de gran ayuda en áreas como la banca (detección de fraudes, análisis de morosidad o segmentación del mercado), telecomunicaciones (control de fugas de clientes, control de redes, ventas cruzadas), seguros (riesgos, mercadeo) y comercial.

Actualmente hay un número creciente de organizaciones inmersas en proyectos de minería de datos. La tecnología se puede aplicar a cualquier organización que disponga de una gran cantidad de datos y que se plantee explotarlos para obtener reglas de negocio o mejorar el servicio que presta.

Se presentan a continuación algunos ejemplos **[KDNU02]**:

- Predicción automática de tendencias y comportamientos:
 - Marketing dirigido: analizar datos sobre envíos por correo publicitarios para identificar el segmento más apropiado para realizar un nuevo mailing.
 - Comportamiento del cliente en supermercados en ciertos días de la semana, de forma que se puedan promocionar ciertos productos en fechas determinadas.
 - Análisis de las ventas de una compañía farmacéutica para reforzar las acciones de marketing en los hospitales y médicos de mayor impacto.
 - Identificación de mejores clientes para el lanzamiento de una nueva tarjeta de crédito.
 - Detección de fraudes en distribuidores de una empresa multinacional.

- Descubrimiento automático de patrones ocultos:
 - Análisis de datos de ventas de productos para identificar aquellos que sin estar relacionados entre sí, se compran juntos.
 - Detección de transacciones fraudulentas con tarjeta de crédito
 - Detección de errores de grabación de datos.
 - Búsqueda de tendencias en la bolsa.
 - Determinación de las causas que producen los fallos en sistemas de producción.
 - Descriptores que “expliquen” los fallos de calidad en el producto.

- Prospectiva:
 - Conseguir modelos aplicables a bases de datos para la selección priorizada de nuevos clientes.
 - Estudios de respuesta ante un posible cambio de precios.
 - Generar nuevos modelos de control de un proceso.

- Segmentación y Clustering:
 - Dividir la base de datos de clientes en segmentos relativamente homogéneos basados en conductas estudiadas.
 - Una organización bancaria puede estudiar qué grupo de usuarios tiene una alta probabilidad de cancelar su cuenta en función de determinados parámetros y a continuación realizar acciones específicas para evitar que ocurra.
 - Clasificar los tipos de clientes de una empresa de seguros.

- Aplicaciones científicas:
 - Análisis de los datos obtenidos a partir de instrumental científico. Esto permite el análisis de los datos para investigación, la formación de hipótesis y teorías.

- Aplicaciones en Biología y Medicina, bioinformación que se traduce en minería de bases de datos distribuidas (por ejemplo el proyecto Genoma).

Industria o Campo	%
CRM / Análisis de Consumidores	25%
Banca	19%
Biotecnología	17%
Seguridad / Anti-terrorismo	16%
Salud	14%
Retail	14%
Política Social/ Análisis de sondeos	13%
Scoring crediticio	13%
Detección de fraude	12%
Medicina / Farmacia	12%
Educación	11%
Marketing Directo / Fundraising	11%
Seguros	11%
Uso de Minería Web	10%
Telecomunicaciones / Cable	9%
Ciencias	7%
e-Commerce	7%
Buscadores / Contenido de web mining	7%
Finanzas	7%
Viajes / Hospedajes	7%
Propaganda por email / Anti-spam	7%
Otros	5%
Gobierno / Militares	5%
Industria	5%
Redes sociales	4%
Entretenimiento / Música / TV / Películas	4%
Publicidad	2%
Software	2%
Inversiones / Stocks	1%

Cuadro 2.3: Encuesta sobre aplicación de la minería de datos
(Fuente: kdnuggets.com, Elaboración: Propia)

2.3.4. DIFICULTADES EN LA IMPLEMENTACION DE LA MINERIA DE DATOS

Se enumeran a continuación algunos de los problemas más habituales a los que se enfrenta cualquier proyecto de Minería de Datos:

- Uno de los mayores problemas es que el número de posibles relaciones es demasiado grande, y resulta prácticamente imposible validar cada una de ellas. Para resolver este problema, se utilizan estrategias de búsqueda, extraídas del área de aprendizaje automático.
- Todas estas herramientas siguen funcionando mejor fijándoles objetivos de búsqueda concretos. Si bien la minería de datos da la impresión de que se puede simplemente aplicar como herramienta a los datos, se debe tener un objetivo, o al menos una idea general de lo que busca.
- El costo de esta prospección de datos debe ser coherente con el beneficio esperado.
- Si bien las herramientas (hardware y software) han bajado su precio, el costo en tiempo, personal y consultoría se ha incrementado, llegando en algunos casos a hacer inviable el proyecto.
- Suele funcionar mejor en problemas ligados a empresas de éxito que en otros casos, debido a la gran dependencia que estas herramientas tienen respecto a todos los estamentos de la empresa, desde mantenimiento a compras.
- Es necesario trabajar en estrecha colaboración con expertos en el negocio para definir modelos. Su ausencia y/o disponibilidad marca el proyecto.
- Otro problema es que la información muchas veces está corrompida, tiene ruido, o simplemente le faltan partes. Para esto, se aplican técnicas estadísticas que ayudan a estimar la confiabilidad de las relaciones halladas.

CAPÍTULO III

PROCESO DE TOMA DE DECISIONES

3.1. OBJETIVO

En la actualidad todas las empresas de telecomunicaciones se ven afectadas por las diversas tipologías de fraude cometido por fraudulentos locales como extranjeros, con el desarrollo de las tecnologías en las telecomunicaciones se han brindado nuevos servicios a los usuarios finales, sin embargo, las nuevas tecnologías también son usadas por los fraudulentos para poder usufructuar con los servicios que ofrece la telefonía. Para poder reducir este impacto las empresas de telecomunicaciones cuentan con áreas encargadas de realizar la lucha frontal contra el fraude en sus diversas modalidades, como principales herramientas cuentan con el Know How que los analistas de fraude han adquirido con el tiempo y los diversos casos resueltos; en cuestiones informáticas se utilizan desde hojas de cálculo, desarrollos InHouse hasta Sistemas Antifraude, todo dependerá del volumen de tráfico cursado y cantidad de analistas asignados para esta tarea, además sin olvidar de la inversión y presupuesto que asignan las empresas de telecomunicaciones en sus respectivas áreas de lucha contra el fraude.

Actualmente los fraudes en las empresas de telecomunicaciones directamente afectan la parte económica e imagen, estos son principalmente el fraude por suscripción, el hurto de línea y la manipulación de la central PBX, para esto el área de lucha contra el fraude es el responsable de mitigar los impactos mediante acciones preventivas, correctivas o disuasivas que dependerán de cada una de las diferentes tipologías de fraude indicadas.

El objetivo principal es implementar un proyecto para optimizar la generación de alertas para la detección de casos de hurto de línea mediante el descubrimiento de patrones utilizando para esto la técnica de la minería de datos evaluando los diversos algoritmos existentes. De esta manera se reducirá considerablemente las pérdidas por fraude, los tiempos de detección, la cantidad de casos de falsos positivos y se elevará la cantidad de casos tipificados como hurto de línea.

3.1.1. OBJETIVO GENERAL

- Reducir la pérdida promedio generada por los casos de hurto de línea.
- Disminuir la cantidad de reclamos por desconocimiento de tráfico.

3.1.2. OBJETIVO ESPECIFICO

- Incrementar en mínimo un 50% la efectividad en la generación de alertas para la detección de hurto de línea.
- Reducir en menos del 20% los casos que son enviados a inspecciones técnicas y que terminan siendo falsos positivos.

3.2. IDENTIFICACIÓN DE PROBLEMA

Lucha contra el fraude (LCF es el área de **tgestión** directamente implicada) ha realizado una tenaz tarea por disminuir el fraude por suscripción (que era su principal problema años atrás) aplicando barreras de entrada al proceso de venta (aplicar requisitos como no contar con deuda vencida e incluir nuevos filtros para validar la titularidad del cliente), implementado nuevos métodos de detección preventiva (se diseño e implemento el monitoreo de alarmas previo a la instalación) y controlando el consumo anómalo en los diversos servicios telefónicos (se diseño e implemento el monitoreo de consumo anómalo internacional). Con las acciones tomadas se redujo la cantidad de casos de fraude por suscripción detectados, debido a que muchos de los casos los pedidos de alta nueva para los servicios telefónicos que solicitaban los fraudulentos, no pasaban los filtros de validación de titularidad. Sin embargo al verse afectados los fraudulentos han migrado sus tácticas de solicitar servicios telefónicos por medios no presenciales (vía Call Center o por agencia) hacia poder vulnerar las facilidades técnicas de la planta instalada y realizar el hurto de línea a diversos servicios telefónicos en diversas zonas de Lima y provincias. Este fraude afecta directamente en pérdidas por ajustes por fraude, número de reclamos de los clientes que desconocen el tráfico facturado y en la manipulación de las facilidades técnicas por terceros que afecta directamente la calidad del servicio.

A continuación se muestra la gráfica donde se observa como luego de la implementación de las barreras de entrada, y los diversos monitoreos que implemento LCF, los casos de fraude por suscripción empiezan a disminuir, sin embargo los casos de hurto de línea empiezan a incrementarse, lo cual lanzo una alerta para empezar a diseñar una estrategia para disminuir los casos afectados por hurto de línea y poder detectar oportunamente los casos y evitar elevadas pérdidas.

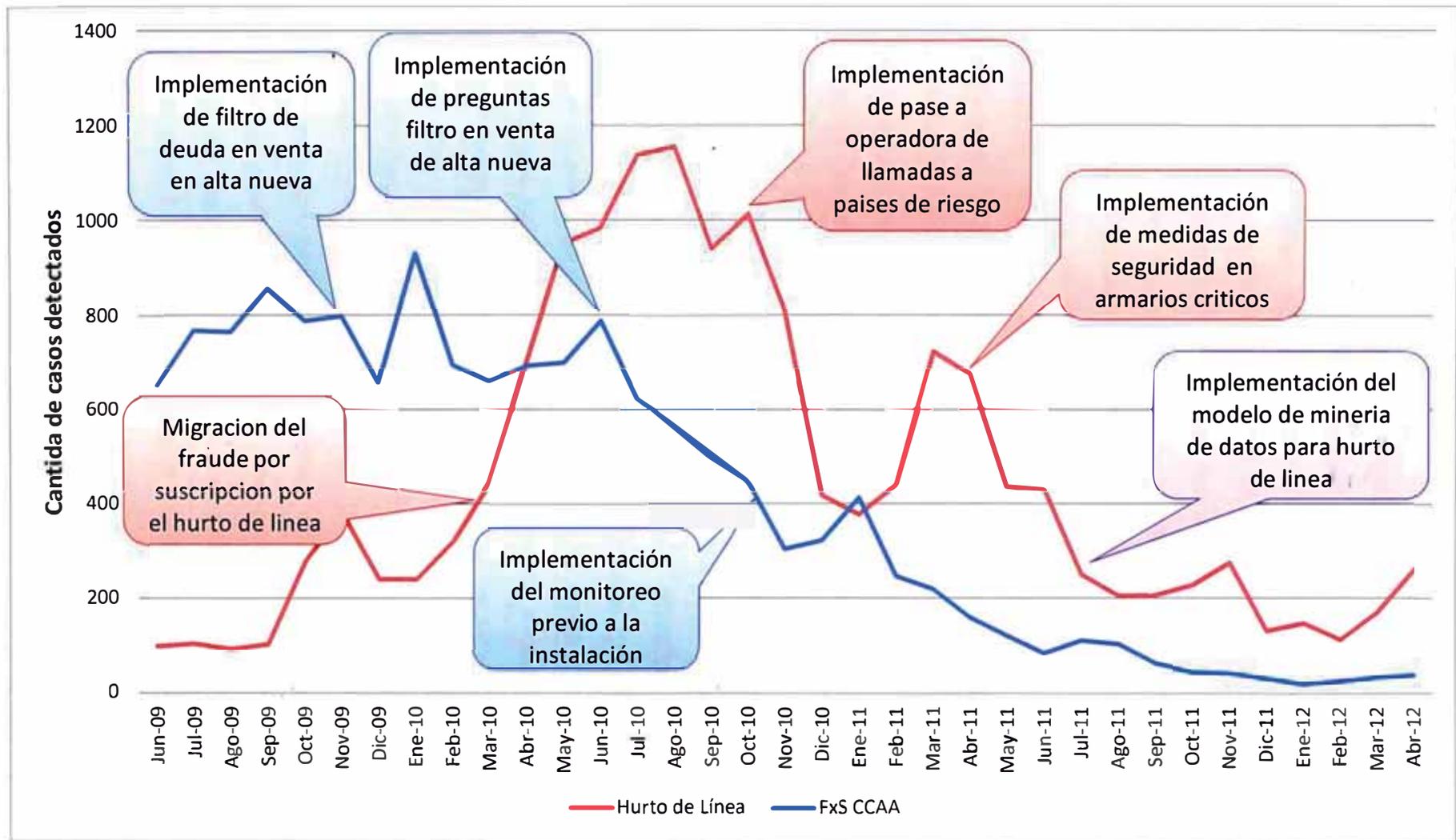


Gráfico 3.1: Infografía de evolutivo de fraude por suscripción vs hurto de línea (Fuente: LCF, Elaboración: Propia)

Actualmente LCF detecta el hurto de línea mediante el monitoreo de consumos anómalos en el tráfico saliente de los diversos servicios telefónicos, sin embargo de esta manera solo se pueden detectar aquellos que realizan un elevado consumo de tráfico internacional (LDI) o celular (PQL), escapando del alcance los servicios telefónicos que realizan hurto de línea de bajo perfil (pocos minutos, pocos destinos, pocas llamadas), también aquellos que le realizan hurto de línea para ser utilizadas en locutorios clandestinos (perfil de reventa de llamadas con pagos regulares). Además que el monitoreo de consumos anómalos no solo detecta hurto de línea, sino también fraude por suscripción, hackeo de PBX, arbitraje o ilegal landing. En la asignación de las alertas que presentan elevados consumos de tráfico internacional y/o celular, no se considera si los destinos ya fueron llamados anteriormente, si es el perfil de llamadas esta dentro de su histórico regular o si las llamadas son realizadas en horarios habituales, etc. con lo cual es más difícil poder identificar los casos de hurto de línea para tomar las acciones correspondientes. Luego de analizar las alertas, aquellas que están tipificadas como hurto de línea se procede a bloquear los tráficos PQL, LDI y LDN para evitar se incremente la pérdida, se envía una solicitud a GORAC para que se realice una inspección técnica a cada uno de los casos, sin embargo debido a la elevada cantidad de casos que recibe GORAC (no solo atiende casos de LCF), no se puede dar abasto y demora en atender una solicitud entre 2 a 6 días, con lo cual existen dos posibles posibilidades, la primera posibilidad que el cliente se dé cuenta que esta sin servicio y presente vía telefónica un reclamo por una avería, con lo cual se envía a una empresa colaboradora, la cual repara la avería, pero no reporta el hurto de línea debido a que existe una penalización porque ellas son responsable de la planta externa; la otra posibilidad es que al bloquear el tráfico saliente del servicio telefónico ya no le es útil al hurtador de línea, el cual vulnerara las facilidades técnicas de otro servicio telefónico dejando libre el servicio telefónico inicial.

En resumen se necesita reducir el impacto económico, para esto se debe detectar oportunamente los casos de hurto de línea, con lo cual se debe de contar con un proceso de generación de alertas efectivas y oportunas, de esta manera se bloquean los servicios telefónicos que han sido realmente afectados y se reduce la cantidad de solicitudes de inspecciones técnicas.

3.3. PLANTEAMIENTO DE ALTERNATIVAS DE SOLUCIÓN.

LCF buscando la mejor estrategia para poder reducir la cantidad de casos de hurto de línea que se presentan en la planta externa, ha detectado que el principal punto a resolver es la generación de las alertas, para lo cual ha presentado las siguientes dos alternativas:

3.3.1. ALTERNATIVA 1: MEJORA DEL PROCESO DE GENERACIÓN DE ALERTAS DE HURTO DE LÍNEA.

Se realizará una mejora en el proceso de generación de alertas rediseñando el perfil de consumos anómalos para detectar específicamente el hurto de línea, diseñando un perfil de cliente, con su consumo de tráfico local, celular, nacional e internacional de los últimos 15 días, de esta manera tener en cuenta si el consumo de llamadas presenta un perfil anómalo o esta fuera del patrón de consumo normal, si presenta llamadas a países de riesgo, llamadas en horarios no acostumbrados, llamadas de larga duración fuera de su histórico de llamadas, llamadas a números marcados en las listas negras de fraude, entre otras variables para ser considerado un posible caso de hurto de línea para analizar.

3.3.1.1. Ventajas:

- Analistas cuentan con el know how necesario para la implementación.
- Presenta un costo reducido de implementación.
- Presenta un tiempo reducido de implementación.
- Reduce la pérdida económica por caso.

3.3.1.2. Desventajas:

- Se necesita contar con nuevos servidores (Hardware) para el procesamiento de la información.
- Presenta una efectividad media en las alertas generadas.
- Presenta una cantidad media de casos falsos positivos,

3.3.2. ALTERNATIVA 2: MODELAMIENTO DEL HURTO DE LÍNEA MEDIANTE LA TÉCNICA DE MINERÍA DE DATOS

Debido al gran volumen de información que se logra recolectar diariamente de los diversos sistemas de la Empresa y que no se puede lograr analizar, ni relacionar en su totalidad, se implementará un modelo de generación de alertas para el hurto de línea, mediante la técnica de Minería de Datos, utilizando la metodología de CRISP – DM, que buscará entender la estructura y el comportamiento de los datos con que se cuentan, transformando los datos para ser probados en los diversos modelos de clasificación y analizar su comportamiento y efectividad, además de realizar mejoras para el rendimiento de los modelos analizados y elegir el modelo que nos brinde una mayor efectividad en la generación de alarmas de posible hurto de línea.

3.3.2.1. Ventajas:

- El modelo de generación de alarmas puede ser adaptable a nuevos servicios.
- Presenta una efectividad elevada en las alertas generadas.
- Presenta un mayor valor agregado al servicio brindado.
- Reduce la pérdida económica por caso.

3.3.2.2. Desventajas:

- Presenta un costo y tiempo de implementación elevado.
- Se necesita capacitación en la técnica de minería de datos.

3.4. SELECCIÓN DE UNA ALTERNATIVA DE SOLUCIÓN.

Para una selección idónea de la mejor alternativa se han tomado en cuenta una serie de criterios para evaluar ambas alternativas, estos criterios han sido desarrollados por un equipo de expertos en el tema de detección y control de fraude.

Los criterios que se han considerado tendrán una ponderación de 1 a 5 sobre el cumplimiento del criterio, siendo 1: Muy Bajo, 2: Bajo, 3: Medio, 4: Alto y 5: Muy Alto. Se considerará para cada uno de los criterios un % entre 0 y 100, para ponderar la importancia de cada criterio.

Los criterios a considerar son los siguientes:

1. **Efectividad de las alertas:** Porcentaje de alertas que se generan para la asignación con perfil de posible hurto de línea y que luego de analizarlas y de realizar la inspección técnica se logra confirmar que existió un caso de hurto de línea en algún punto de las facilidades técnicas.
2. **Adaptabilidad para nuevos servicios:** La alternativa elegida permitirá adaptarse para poder brindar el servicio a nuevos clientes en otros segmentos del mercado.
3. **Menor costo de implementación:** Monto que se invertirá para desarrollar la solución, considerando desde capacitación, compra de hardware, compra de licencias de software, implementación, mantenimiento, etc.
4. **Menor tiempo de implementación:** Cantidad de semanas que se demorará desde el inicio de puesta en marcha de la alternativa hasta su implementación de la misma.
5. **Menor cantidad de casos Falsos Positivos:** Porcentaje de alertas que se asignan con perfil de posible hurto de línea y que al realizar la inspección técnica oportunamente se valida de que no existió hurto de línea.
6. **Reducción en las pérdidas económicas:** Los casos de hurto de línea tiene una pérdida económica por el tráfico cursado, que luego el cliente lo reclamará desconociéndolo o por la reparación de las facilidades técnicas que fueron manipuladas y alteradas por un tercero.

3.4.1. CUADRO DE EVALUACION DE CRITERIOS VS ALTERNATIVAS

De los criterios elegidos por el grupo de expertos se procedió a darles un nivel de importancia (%) y se pondero, obteniendo como resultado el siguiente cuadro:

Criterios	% Importancia	ALTERNATIVAS			
		MEJORA DEL PROCESO DE GENERACIÓN DE ALERTAS DE HURTO DE LÍNEA		MODELAMIENTO DEL HURTO DE LÍNEA MEDIANTE LA TÉCNICA DE MINERÍA DE DATOS	
		Puntaje	Ponderado	Puntaje	Ponderado
Efectividad de las alertas	25	2	0.5	4	1
Adaptabilidad para nuevos servicios	20	2	0.4	4	0.8
Menor costo de implementación	18	4	0.72	1	0.18
Menor tiempo de implementación	15	4	0.6	2	0.3
Menor cantidad de casos Falsos Positivos	13	2	0.26	3	0.39
Disminución en las pérdidas económicas	9	3	0.27	4	0.36
Totales	100	17	2.75	18	3.03

Cuadro 3.1: Tabla de evaluación de criterios (Elaboración propia)

Luego de la evaluación, se concluye que la alternativa elegida es la Alternativa 2: Modelamiento del hurto de línea mediante la técnica de Minería de Datos.

3.5. PLANES DE ACCIÓN PARA DESARROLLAR LA SOLUCIÓN ELEGIDA.

Para la solución elegida, se tomara como metodología de desarrollo la CRIPS-DM (CRoss Industry Standard Process for Data Mining), que consta de 6 fases relacionadas entre sí de una forma cíclica, es decir con retroalimentación, cada una de las cuales cuenta con sus propias actividades y a la vez estas con sus propios entregables.

Etapa	Actividad	Entregables
FASE I: Comprensión del Negocio	Determinar objetivos del negocio	Escenario actual
		Objetivos del negocio
		Factores críticos para el éxito del negocio
	Evaluación de la situación	Inventario de recursos
		Riesgos y Contingencias

		Terminología
		Costos y beneficios
	Determinar los objetivos de minería de datos	Objetivos de minería de datos
		Factores críticos de éxito para minería de datos
	Desarrollar el plan del proyecto	Plan del proyecto
FASE II: Comprensión de los datos	Recolección inicial de datos	Reporte de la colección de datos
	Descripción de los datos	Reporte con la descripción de los datos
	Exploración de los datos	Reporte con los resultados de la exploración de los datos
	Verificación de la calidad de los datos	Reporte con los resultados de la calidad de los datos
FASE III: Preparación de los datos	Preparación de la colección de datos	Colección de datos
		Descripción de la colección de datos
	Selección de los datos	Criterio de Inclusión/Exclusión de datos
	Limpieza de datos	Reporte de limpieza de datos
	Construcción de los datos	Atributos derivados
		Generación de registros
Integración de los datos	Unificación de los datos	
Formateo de los datos	Reformateo de los datos	
FASE IV: Modelado	Seleccionar la técnica de modelado	Técnica de modelado
		Supuesto de modelado
	Generar diseño de pruebas	Diseño de pruebas
	Construir el modelo	Configuración de parámetros
		Modelos
		Descripción del modelo
Evaluar el modelo	Evaluación del modelo	
	Revisión de los parámetros de configuración	
FASE V: Evaluación	Evaluación de resultados	Evaluación de los resultados del proceso de minería de datos
		Aprobación del modelo
	Proceso de revisión	Revisión del proceso
	Determinar los próximos pasos	Lista de posibles acciones
Decisión a tomar		
FASE VI: Implementación	Plan de implementación	Plan de implementación
	Plan de monitoreo y mantenimiento	Plan de monitoreo y mantenimiento
	Armado del informe final	Informe final
		Presentación final
Revisión del proyecto	Documentación de la experiencia	

Cuadro 3.2: Fases, actividades y entregables de la metodología CRISP-DM (Fuente: kdnuggets.com, Elaboración: Propia)

La metodología CRISP-DM en el nivel superior se divide en 6 fases:

Fase I: Entender el Negocio;

Esta fase se enfoca en la comprensión de los objetivos de proyecto y exigencias desde una perspectiva de negocio, en la definición de un problema de minería de datos y en un plan preliminar diseñado para alcanzar los objetivos.

Fase II: Entender los Datos:

Esta fase pretende coleccionar los datos iniciales, desarrollar procedimientos para entender los datos, identificar los problemas de calidad de datos, descubrir los primeros conocimientos en los datos; y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.

Fase III: Preparación de los datos:

Esta fase cubre todas las actividades necesarias para construir el conjunto de datos que serán entregados a las herramientas de modelado desde los datos en brutos iniciales. Las tareas de preparación de datos probablemente se ejecutaran muchas veces y en cualquier orden. Las tareas incluyen la selección de tablas, registros, y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

Fase IV: Modelado:

En esta fase, varias técnicas de modelado son aplicadas y calibradas. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de datos. Esta fase se relaciona con la fase de evaluación, todo modelo debe ser evaluado para calcular su rendimiento.

Fase V: Evaluación:

En esta fase ya se ha construido un modelo óptimo que tiene la mejor calidad de una perspectiva de análisis de datos.

Fase VI: Desarrollo:

En esta fase el modelo desarrollado se aplica para resolver el problema planteado inicialmente.

Se diseño el siguiente cronograma para el desarrollo e implementación del proyecto de minería de datos para la generación de alertas de hurto de línea según se muestra a continuación:

Actividades Principales	Mes1		Mes2		Mes3		Mes4		Mes5		Mes6		Mes7																
	Se m1	Se m2	Se m3	Se m4	Se m5	Se m6	Se m7	Se m8	Se m9	Se m10	Se m11	Se m12	Se m13	Se m14	Se m15	Se m16	Se m17	Se m18	Se m19	Se m20	Se m21	Se m22	Se m23	Se m24	Se m25	Se m26	Se m27	Se m28	
Entender el Negocio	■																												
Entender los Datos			■	■																									
Preparación de los datos					■	■	■	■	■																				
Modelado													■	■	■														
Evaluación																							■	■	■				
Desarrollo																												■	■

Cuadro 3.2: Fases, actividades y entregables de la metodología CRISP-DM (Fuente: LCF, Elaboración: Propia)

Describiremos con detalle cada una de las fases y sus actividades respectivas para la solución elegida.

3.5.1. FASE I: COMPRENSIÓN DEL NEGOCIO

La metodología de CRISP-DM permite definir todo el proceso de descubrimiento de conocimiento en bases de datos, definiendo fases y tareas y el significado de cada una de ellas. Lógicamente, la primera de estas fases corresponde con la comprensión del negocio y el análisis del problema, donde se busca garantizar la perfecta

comprensión del problema planteado y poder llegar así a una definición lo más completa posible de los objetivos finales.

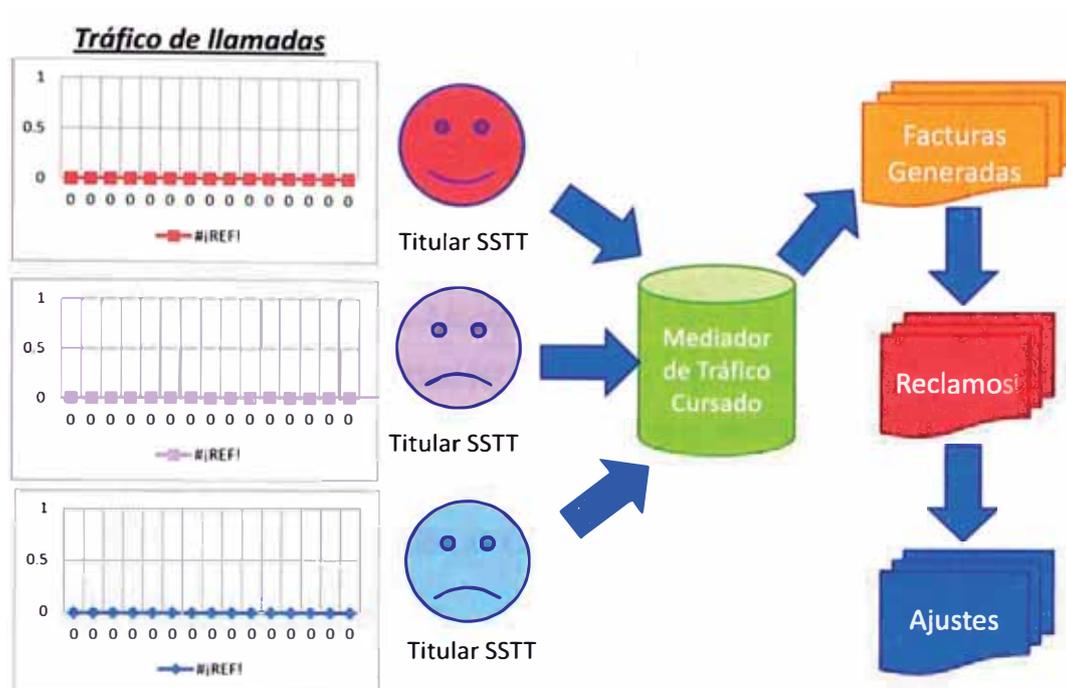


Gráfico 3.2: Representación básica del Negocio (Fuente y Elaboración: Propia)

3.5.1.1. Determinar objetivos del Negocio

Primeramente se deben especificar claramente el problema que sustenta la realización del proyecto junto con los objetivos de negocio esperados. Esta actividad es fundamental, ya que un error en la definición de los objetivos se arrastrará en todos los procesos posteriores. Las fases fundamentales son:

3.5.1.1.1. Escenario Actual:

Se busca recoger toda la información disponible sobre el entorno en el que se desarrolla el proyecto al comienzo del

mismo, con el fin de identificar los objetivos y los recursos disponibles.

- Organización:
 - Unidades implicadas:
 - Supervisión Red Fija - tgestiona
 - Unidades Afectadas:
 - Supervisión Red Fija - tgestiona
 - Dirección de Riesgo Operacional - TdP
 - GORAC – TdP
 - Atención al Cliente - TdP

- Problema:
 - Unidad del negocio directamente afectada:
Supervisión de Red Fija de la Jefatura de Lucha contra el fraude y PIE.

 - Descripción del problema:
El problema ha sido descrito detalladamente en el punto 3.2 Identificación del Problema, sin embargo, en resumen se necesita incrementar la efectividad de las alertas, se debe detectar oportunamente los casos de hurto de línea, para lo cual se debe de contar con un proceso de generación de alertas efectivas y oportunas, para que de esta manera se bloqueen los servicios telefónicos que han sido realmente afectados y se reduce la cantidad de solicitudes de inspecciones técnicas.

- **Necesidades y Expectativas:**

Principalmente es mejorar la efectividad en la generación de las alertas que cuentan con perfil de hurto de línea.

- **Equipo de trabajo:**

Está formado por un reducido grupo de colaboradores sensibilizado y conocedor del proceso del negocio, lo que simplifica la transmisión de conceptos propios de la minería de datos, además de haber sido capacitados en minería de datos para un mejor desarrollo del proyecto.

3.5.1.1.2. *Objetivos de Negocio:*

Los objetivos principales son:

- Incrementar la efectividad en la generación de alertas para la detección de hurto de línea.
- Reducir la cantidad de casos de enviados a inspecciones técnicas que terminan siendo falsos positivos.
- Reducir la pérdida generada por los casos de hurto de línea.
- Disminuir la cantidad de reclamos por desconocimiento de tráfico.

3.5.1.1.3. *Criterios de Éxito:*

El grado de cumplimiento de los objetivos anteriormente descritos será medido por los siguientes criterios:

- % de efectividad de las alertas generadas para la asignación y análisis respectivo.

- % de casos que luego de la inspección técnica se indico que no existieron evidencias de hurto de línea.
- Monto mensual de pérdida promedio por tráfico generado en los casos que fueron tipificados como hurto de línea.
- Numero de reclamos mensual de clientes que desconocen haber realizado un tráfico facturado.

3.5.1.2. Evaluación de la situación

En esta segunda actividad se analiza con más detalle todos los factores que pueden influir y afectar al desarrollo posterior del proceso de la minería de datos. Las fases en que se divide son:

3.5.1.2.1. Recursos Disponibles:

A continuación se realiza una descripción de los recursos disponibles:

- Fuentes de Datos:

Los sistemas de adquisición de datos son los siguientes:

- DataWareHouse TdP
- Datamart ATIS
- Proceso de Fraude
- SAMARA
- FENIX

- Hardware:

Para el análisis y desarrollo de las técnicas de minería de datos se cuenta con:

- Tres PCs HP con procesadores Core2Duo de 2.1 GHz, 4 GB de memoria RAM, 320 GB de disco duro, además cuenta con sistema operativo Windows XP.
 - Una PC HP con procesador QuadCore de 2.0 GHz, 4 GB de memoria RAM, 500 GB de disco duro, además cuenta con sistema operativo Windows XP, esta PC se utilizará para el entrenamiento de los modelos.
- **Software:**

Se disponen del siguiente software, para diversas tareas:

- SQL Server Profesional 2008
- PHP 5
- Herramienta GNU de Minería de datos: WEKA
- RAPID MINER
- Microsoft Office: Word, Excel, PowerPoint, etc.
- Sistema Operativo: Windows XP
- Add-In de Minería de Datos para MS Excel

3.5.1.2.2. *Requerimientos, Supuestos y Restricciones:*

Se realiza una lista de requerimientos, restricciones y supuestos relativos tanto a la planificación del proyecto como a los datos o recursos disponibles.

- Debido a temas de confidencialidad, algunos datos deberán ser alterados o modificados.
- Se necesita contar la total disponibilidad del personal asignado al proyecto, hasta que se logre implementar el mismo.

- Se necesita contar con data histórica que permita generar un modelo correcto.
- Se necesita contar con casos históricos de hurto de línea, casos históricos de falsos positivos en hurto de línea, para poder entrenar y probar el modelo.
- Los resultados de los modelos deberán ser revisados por los expertos de la supervisión.
- Se llevara un cronograma del proyecto.
- La comunicación será tanto verbal en reuniones, como escrita vía email para los temas diversos.

3.5.1.2.3. Riesgos y Contingencias:

Se realizará una lista de circunstancias que puedan retrasar o impedir la realización del proyecto y se planificarán las acciones a llevar a cabo si se producen.

- Debido a la inexperiencia del equipo en otros proyectos de minería de datos, pueden existir algunos puntos muertos, que deberían buscar una rápida solución para evitar estancamientos.
- Se necesita una comunicación fluida y flexible entre el equipo del proyecto y las partes interesadas. Los requerimientos deben ser claramente indicados.
- Se debe revisar la calidad de la data generada por los diversos sistemas para evitar inconsistencias puntuales o generales. Además de contar con la data en el momento oportuno y necesario.
- Las PCs utilizadas deben de tener un alto performance para poder procesar la gran cantidad de información en tiempos reducidos.

3.5.1.2.4. Costos y Beneficios:

Los costos planteados principalmente corresponden a los relativos a la capacitación en técnicas de minería de datos, al personal encargado, del hardware y de las licencias del software para el procesamiento de la información.

En los beneficios se deberán considerar lo siguiente:

- Reducción de la deuda promedio por caso detectado de hurto de línea.
- Reducción del número promedio de reclamos por desconocimiento de tráfico.
- Reducción del número de inspecciones técnicas que terminan siendo falso positivo.

3.5.1.3. Determinar los objetivos de minería de datos

Luego de plantear los objetivos del negocio y realizar un inventario de los recursos humanos y no humanos que disponemos, la metodología CRISP-DM indica que debemos desarrollar los objetivos desde una perspectiva más técnica, se pretende adaptar los objetivos planteados desde el punto de vista de la técnica de minería de datos.

A continuación se presentan en orden de desarrollo, los objetivos del proceso de la minería de datos, así como los criterios para valorar el grado de éxito conseguido.

3.5.1.3.1. Objetivos del Proyecto de minería de datos

- Selección de las variables más importantes del proceso y su grado de importancia, mediante las técnicas estadísticas y de minería de datos.
- Generación de conocimiento mediante las técnicas para determinar los patrones de comportamiento de los diversos sucesos que pueden mejorar la generación de alertas con perfil de hurto de línea.
- Generación de reglas que permitan modelar categóricamente las variables y los sucesos diversos.
- Generación y elección de un modelo de minería de datos, que tenga mayor efectividad en la generación de alertas con perfil de hurto de línea.
- Desarrollo de una interfaz grafica para poder explotar el modelo elegido.

3.5.1.3.2. Criterios de éxito del Proyecto de minería de datos

Para verificar los objetivos anteriores, es necesario establecer los criterios que determinen el grado de éxito de los resultados anteriores. De esta forma, para cada uno de los puntos anteriores se formulan los siguientes factores de éxito:

- Para la selección de las variables más importantes del proceso:
 - Grado de eficacia de los modelos obtenidos con esas variables.
 - Estimación del grado de confiabilidad de los expertos.
- Para la generación del conocimiento de las causas que puedan mejorar la generación de alertas:

- Valoración de los expertos del conocimiento adquirido.
 - Mejora de la generación de alertas debida a este conocimiento.
 - Uso que se realiza de los conocimientos generados.
- Para la generación de las reglas:
 - Grado de uso que realizan los expertos de las mismas.
 - Mejora del proceso debido al uso de estas reglas.
 - Para el modelo generado:
 - Porcentaje de alertas que no cuentan con perfil de hurto de línea.
 - Porcentaje de alertas tipificadas como hurto de línea que al realizar la inspección técnica, no se encontraron evidencias.
 - Para la interfaz grafica:
 - Aprobación por parte de los expertos.
 - Experiencia de uso del analista.

3.5.1.4. Desarrollar el plan del proyecto

Por último, dentro de la primera fase de la metodología CRISP-DM, se procede a realizar la planificación inicial teórica de las tareas a realizar y de las herramientas a usar.

3.5.1.4.1. Planificación del proyecto y técnicas propuestas:

Dentro de la planificación del proyecto se plantearon las siguientes tareas:

- Desarrollo de la lista inicial de variables a tratar a partir del conocimiento de los expertos y de la documentación del proceso actual.
 - Tiempo Previsto: 2 semanas.
 - Métodos y Técnicas Previstas: Delphi, contraste de opiniones, estudio de documentación.
 - Recursos Necesarios Previstos:
 - Expertos.
 - Documentación del proceso que describa el modelo actual.
 - Microsoft Excel 2007.
 - Entrada: las diversas variables de los sistemas.
 - Salida Prevista: Primera lista de variables a trabajar.

- Recolección de los datos de las base de datos diversas.
 - Tiempo Previsto: 4 semanas.
 - Métodos y Técnicas Previstas: Tratamiento con herramientas de base de datos.
 - Recursos Necesarios Previstos:
 - MySQL.
 - DataWareHouse y Datamart.
 - Microsoft SQL 2008.
 - PC Core2Duo de 4 GB memoria RAM.
 - Entrada: Tráfico diario de los servicios telefónicos.
 - Salida Prevista: Base de Datos perfilada por los servicios telefónicos

- Exploración de los datos mediante técnicas estadísticas y de visualización multivariante, selección de variables más características, estudio y filtrado de espurios, reducción dimensional, transformación de los datos:
 - Tiempo Previsto: 10 semanas.
 - Métodos y Técnicas Previstas: Técnicas de visualización estadística y multivariante.
 - Recursos Necesarios Previstos:
 - Herramienta R de análisis estadístico
 - RAPID MINER.
 - Microsoft Excel 2007.
 - PC Core2Duo de 4 GB memoria RAM.
 - Entrada: Base de Datos obtenida en el paso 2.
 - Salida Prevista: Base de datos filtrada, con las variables más importantes, libre de ruidos y con los datos transformados.

Si los datos no son adecuados, será necesario volver al punto 2. Si se supone necesaria alguna variable no considerada inicialmente, habrá que recomenzar en el punto 1.

- Extracción de patrones de comportamiento, reglas, etc.
 - Tiempo Previsto: 5 semanas.
 - Métodos y Técnicas Previstas: Árboles clasificadores, reglas de decisión, reglas asociativas, herramientas de clusterizado y visualización multivariante, etc.
 - Recursos Necesarios Previstos:
 - Microsoft SQL Server Analysis Service.
 - Add-In de Minería de Datos para MS Excel.
 - RAPID MINER.
 - PC Core2Duo de 4 GB memoria RAM.

- Entrada: Base de datos con las variables más importantes, libre de ruidos y con los datos transformados.
- Salida Prevista: Reglas y árboles que expliquen los comportamientos anómalos.

Si los datos no son adecuados, será necesario volver al punto 3.

- Creación del modelo de minería de datos: entrenamiento y validación, búsqueda del modelo óptimo.
 - Tiempo Previsto: 4 semanas.
 - Métodos y Técnicas Previstas: modelos de minería diversos (árboles de decisiones, redes neuronales, etc.).
 - Recursos Necesarios Previstos:
 - Microsoft SQL Server Analysis Service.
 - Add-In de Minería de Datos para MS Excel.
 - RAPID MINER.
 - PC QuadCore de 4 GB memoria RAM.
 - Entrada: Base de datos con las variables más importantes, libre de ruidos y con los datos transformados.
 - Salida Prevista: Modelo de minería de datos con mejor calidad de predicción que el actual modelo.
- Desarrollo de la interfaz gráfica para la explotación del modelo generado.
 - Tiempo Previsto: 3 semanas.
 - Métodos y Técnicas Previstas: Metodología de desarrollo de interfaces gráficas.

- Recursos Necesarios Previstos:
 - PHP 5.
 - Microsoft SQL Server Analysis Service.
 - Add-In de Minería de Datos para MS Excel.
 - PC Core2Duo de 4 GB memoria RAM.
- Entrada: Datos históricos para la fase de simulación. Datos actuales para el proceso de monitorización y ajustes.
- Salida Prevista: Interfaces gráficas para el uso del modelo generado.

3.5.2. FASE II: COMPRENSIÓN DE LOS DATOS

Esta fase comprende el análisis de los datos, donde se adquieren los datos, se describen y se analiza su calidad. Se divide en varias etapas secuenciales:

- Adquisición de los datos.
- Descripción de los datos.
- Exploración de los datos.
- Verificación de la calidad de los mismos.

Este proceso se debe de realizar hasta que la calidad de los datos sea lo suficientemente buena para la consecución de las fases siguientes y que garanticen el éxito.

3.5.2.1. Recolección inicial de datos

Consiste en recolectar los datos de las diferentes fuentes con las que contamos. Necesitamos generar un conjunto inicial de datos con los que trabajaremos luego.

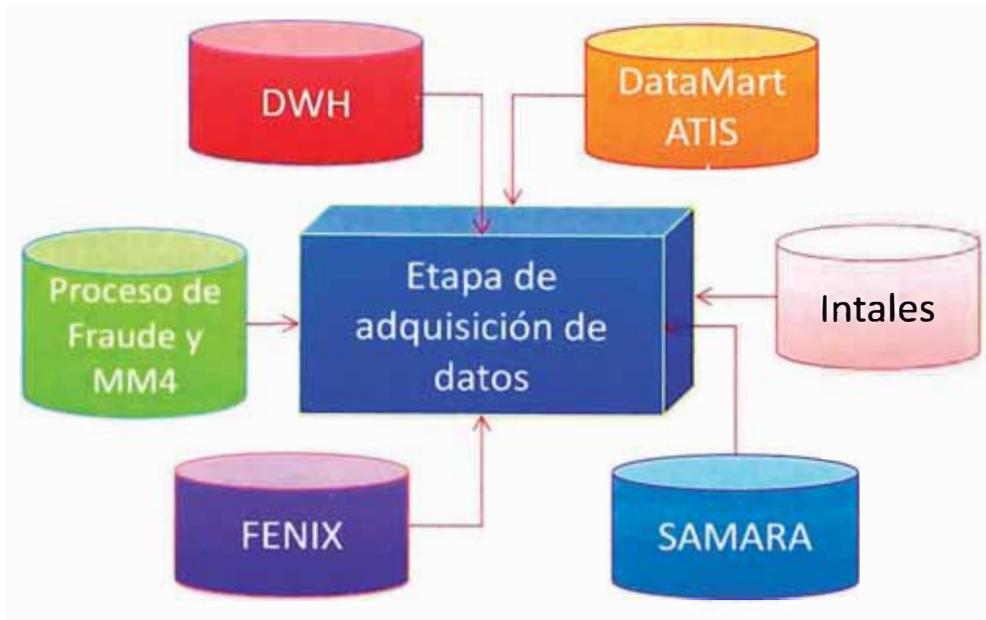


Gráfico 3.3: Fuentes de datos diversas (Fuente: LCF, Elaboración: Propia)

3.5.2.2. Descripción de los datos

En esta etapa se describen las diversas fuentes de datos, que se lograron adquirir en la etapa anterior.

- **DWH:** Se cuenta con todas las facilidades técnicas y data comercial de los servicios telefónicos de toda la planta instalada, además de todo el tráfico histórico de los últimos 6 meses, se cuenta con data con un día de desfase.
- **DataMart ATIS:** Se cuenta con todas las peticiones de venta o postventa que han realizado los servicios telefónicos de toda la

planta instalada, se cuenta con data con un día de desfase. Cuenta con problemas de carga y demora de entrega.

- **Proceso de Fraude y MM4:** Brinda el tráfico acumulado diario y el tráfico detallado de los servicios telefónicos que han realizados llamadas, se cuenta con data con un día de desfase.
- **Intales:** Brinda solamente el tráfico detallado de larga distancia internacional de los servicios telefónicos que hayan realizado este tipo de tráfico, se cuenta con data con dos horas de desfase.
- **FENIX:** Brinda los reclamos registrados por los diversos servicios telefónicos, se cuenta con data en línea.
- **SAMARA:** Brinda todos los casos que han sido analizados por Lucha contra el fraude, nos permitirá contar con los casos que han sido tipificados como hurto de línea, además de los falsos positivos de hurto de línea, se cuenta con data en línea.

3.5.2.3. Exploración de los datos

A partir de la descripción de los datos, se realiza un primer análisis superficial de las características de los datos, como por ejemplo:

- Relaciones entre variables.
- Tipo de distribución de los datos.
- Agrupamientos

Se debe de realizar el análisis de todos los campos de las fuentes de datos elegidas, como paso inicial usaremos el software

estadístico R y algunas herramientas de visualización multivariantes como por ejemplo: gráficos de cajas o box plots, histogramas, rectas cuantil - cuartil, etc.; también nos permitirá realizar una exploración inicial, donde podemos comparar varias variables entre sí mostrándonos en un mismo gráfico el histograma, la correlación y la distribución de los pares de datos de las variables, también utilizaremos el RAPID MINER. Para más detalle ver los anexos.

3.5.2.4. Verificación de la calidad de los datos

En esta etapa, se determino la calidad de los datos disponibles para validar que tan buenos son. Para ello se analizo si: Los datos tienen errores, describen realmente la realidad, presentan campos vacios, presentan ruido existente, existe consistencia entre las fuentes de datos, existen datos redundantes.

Utilizando el software estadístico R y el RAPID MINER, nos permitió analizar las diversas variables y verificar la calidad de los datos, según los requerimientos descritos en el párrafo anterior. Para más detalle ver los anexos.

3.5.3. FASE III: PREPARACIÓN DE LOS DATOS

En esta Fase tratamos de generar una base de datos óptima para la posterior fase de modelado. Esta fase consta de las siguientes etapas:

- Selección de los datos.
- Limpieza de los datos.
- Generación de variables adicionales.

- Integración de orígenes de datos.
- Cambio de formato de los datos.

Estas etapas se realizaron repetidamente hasta que obtuvimos una base de datos lo suficientemente adecuada para las siguientes fases.

3.5.3.1. Selección de los datos

Después de obtener y describir los datos en la fase anterior, se realiza la selección de las variables más representativas, que deben de cumplir con lo siguiente:

- Se busca que sean independientes entre sí.
- Deben de contener datos fiables y sin ruidos.
- Deben de contener relevancia individual destacada.
- Que sean lo más descriptivas para el problema que vamos a analizar.
- Se eliminaran las variables dependientes y aquellas cuya influencia en el problema sea nula o muy escasa.

En la selección de los datos se han eliminado la información del Datamart ATIS y del FENIX debido a que no presentan información relevante para poder resolver el problema,

Se cuenta con tres fuentes de detalle de tráfico (Proceso de Fraude, MM4 e Intales), después del análisis se llegó a determinar que las fuentes necesarias para analizar tráfico LDI es Intales, para el resto de tráfico se utilizó el MM4.

Se ha seleccionado el universo del DataWareHouse que nos brinda las facilidades técnicas de los servicios telefónicos y

SAMARA que nos brindara los casos que han sido analizados y tipificados como alerta y los casos de falso positivo.

3.5.3.2. Limpieza de datos

En esta etapa principalmente se busca que las variables elegidas sean lo más fiables posibles, principalmente se eliminara o corregirá el ruido en los datos y los espurios. Para poder realizar lo antes descrito se utilizaron técnicas estadísticas que nos permiten corregir, ignorar o eliminar los datos con problemas en nuestras variables. A la par que se realizaba la implementación se desarrollaron mejoras para una mejor adquisición de los datos sin ruidos, ni espurios. Para más detalle ver los anexos.

3.5.3.3. Generación de variables adicionales

Esta etapa busca generar nuevas variables a partir de las ya existentes siempre que permitan agilizar el estudio del problema. Se incluye la transformación de los datos para normalizar, asignar nuevos pesos, recodificar las variables, también se puede utilizar indicadores estadísticos y transformadas. Para más detalle ver los anexos.

3.5.3.4. Integración de orígenes de datos

Se combinarán los datos procedentes de diferentes orígenes, siempre que no se haya hecho ya, para obtener una base de datos más compacta y útil. Para más detalle ver los anexos.

3.5.3.5. Cambio de formato de los datos

Esta etapa busca adecuar el formato de los datos para ser usados por las herramientas que se utilicen en fases posteriores, se puede cambiar el tipo de variable, reordenar los datos, etc. Para más detalle ver los anexos.

3.5.4. FASE IV: MODELADO

Del análisis de las fuentes de datos y de las principales variables se han obtenido algunas conclusiones preliminares, en esta etapa se busca extraer información más útil que ayude a comprender la interrelación entre las variables analizadas, buscando conclusiones que nos permitan entender el proceso. Para conseguir lo descrito se aplican diferentes técnicas de minería de datos para obtener información práctica, oculta en la base de datos, que nos permita extraer conclusiones importantes y nos ayuden a resolver nuestras diversas consultas.

3.5.4.1. Seleccionar la técnica de modelado

Para esta etapa se desarrollaron diversos algoritmos como árboles de decisión, regresión lineal, regresión polinómica de grado 4 y redes neuronales, probándose en cada caso su capacidad predictiva, buscando el modelo que brinde mayor efectividad a la solución de nuestro problema.

3.5.4.2. Generar diseño de pruebas

En esta etapa se van a seleccionar el tamaño de la muestra a trabajar todas las pruebas, además de preparar varios conjuntos de datos para las pruebas que contengan diversos porcentajes de casos que fueron hurto de línea (verdadero positivos) y los que después de una inspección técnica se concluyó que no fue hurto

de línea (falso positivo). Se entrenara en las pruebas con el 60% de los datos extraídos de forma aleatoria de los conjuntos de datos, usando otros 20% para la validación y el último 20% para testeo.

3.5.4.3. Construir el modelo

Una vez que se cuentan con los conjuntos de datos elegidos aleatoriamente, aplicaremos cada una de las técnicas de modelado siguientes árboles de decisión, regresión lineal, regresión polinómica de grado 4 y redes neuronales, para poder desarrollar esta evaluación utilizaremos el software WEKA y el RAPID MINER, donde se lograra definir una descripción detallada del modelo, sus parámetros, su exactitud y la sensibilidad. Para más detalle ver los anexos.

3.5.4.4. Evaluar el modelo

Para evaluar los diversos modelos se utilizó la matriz de validación cruzada, donde se miden los % de casos verdaderos positivos (casos positivos predichos por el modelo que resultaron realmente positivos) y el % de casos falsos positivos (casos positivos predichos por el modelo que resultaron realmente negativos)

Como resultado de las pruebas con los diversos modelos, las redes neuronales tuvieron el mejor rendimiento, seguido por los árboles de decisión.

3.5.5. FASE V: EVALUACIÓN

En esta fase del proceso se evalúan los modelos obtenidos, se revisa el proceso desarrollado hasta el momento y se busca la forma de mejorar los resultados obtenidos.

Para poder validar los resultados obtenidos, es necesario preparar un nuevo conjunto de datos que nos permita determinar el grado de eficiencia, trabajando con nueva información.

3.5.5.1. Evaluación de resultados

En esta etapa se busca comprobar que los modelos cumplen los objetivos buscados, así como determinar las flaquezas de los mismos y la forma de reducirlas.

El algoritmo fue optimizado tanto en sus parámetros internos como en parámetros dependientes del método de muestreo de los datos para el proceso de entrenamiento, donde se logró el mejor desempeño para muestras de entrenamiento de 72% de casos de hurtos de línea y 28% de casos de no hurto de línea. Para más detalle ver los anexos.

3.5.5.2. Proceso de revisión

Después de analizar los resultados de la minería de datos y de los modelos, se procede a revisar todo el proceso realizado buscando las alternativas de mejora. Se busca mejorar toda la metodología para llegar a un afinamiento mayor de los resultados.

Con el modelo óptimo creado en la fase anterior se evaluó con nuevos registros no utilizados en la fase de modelado, identificando nuevas relaciones que aportan al modelo. Para más detalle ver los anexos.

3.5.5.3. Determinar los próximos pasos

Si los resultados son los más adecuados, se pasará a la fase de implementación, sino, se desarrollará una serie de alternativas de mejora de las fases previas. De esta forma, según el grado de éxito de los resultados obtenidos, los plazos y recursos de que se disponen y el potencial esperado, se decide la alternativa mejor, justificando su elección.

Se logro determinar que el modelo de redes neuronales elegido estaba listo para ser implementado y explotado.

3.5.6. FASE VI: IMPLEMENTACIÓN

En esta fase ya se ha construido un modelo que tiene la mejor efectividad. La finalidad es usar el modelo óptimo para predecir comportamientos de nuevos registros de datos.

3.5.6.1. Plan de implementación

En esta etapa se busca realizar el despliegue del modelo, donde se automatizará gran parte de las fases del proyecto de minería de datos, creando scripts y procedimientos almacenados para la configuración de parámetros, importación, procesamiento y carga de datos.

Principalmente se desarrollo del aplicativo que explotaría al modelo de redes neuronales, donde se podría de manera amigable realizar la generación de alertas que tienen perfil de hurto de línea.

3.5.6.2. Plan de monitoreo y mantenimiento

Esta etapa prepara un plan que permita monitorizar el rendimiento del modelo de redes neuronales, debido a que el fraude es dinámico y modifica su comportamiento cada vez que es detectado, por esta razón cuando la efectividad del modelo empieza a bajar constantemente se debe de realizar un mantenimiento del modelo, buscando entrenar nuevamente el modelo de redes neuronales para encontrar nuevos patrones y incrementar la efectividad.

3.5.6.3. Armado del informe final

En esta etapa se genera el reporte final donde se expone todo lo desarrollado con la metodología CRISP-DM para nuestro proyecto de minería de datos, además se genera la presentación donde se expondrá todos los pormenores de la experiencia, los pasos realizados, sus resultados u conclusiones obtenidas.

3.5.6.4. Revisión del proyecto

En esta etapa se busca principalmente la documentación de toda la experiencia realizada, para poder ser replicada en el futuro o revisada para nuevos proyectos. Se necesita que se consideren todas las fases, etapas, actividades, etc. y queden correctamente documentadas, que faciliten su posterior consulta.

CAPITULO IV

ANÁLISIS DE RESULTADOS

En este capítulo se evaluará el desarrollo e implementación del proyecto de minería de datos para desarrollar un modelo que incremente la efectividad al generar las alertas de detección de hurto de línea, describiremos los criterios que utilizaremos para evaluar, además de desarrollar el flujo de caja que nos permitirá generar los indicadores económicos necesarios.

4.1. SELECCIÓN DE CRITERIOS DE EVALUACIÓN

Con respecto a los criterios de evaluación para los proyectos informáticos, han ido cambiando simultáneamente con la evolución de la tecnología de la siguiente forma:

- Primeramente, sólo tenía importancia la evaluación costo - beneficio.
- Luego se evaluaba además si se facilitaba la obtención de los objetivos de la organización y si las tecnologías de información mejoraban la calidad de las inversiones.
- Posteriormente cobró importancia cómo las tecnologías de información podían mejorar la toma de decisiones y aumentar la participación de mercado.

- Ahora se da mayor importancia a cómo las tecnologías de información pueden aumentar la capacidad de la información y la innovación.

Un importante elemento para evaluar un proyecto informático, es la evaluación costo - beneficio. Sin embargo, tiene que estar sustentado por beneficios cuantitativos o cualitativos. Es muy importante la coherencia del proyecto con los planes estratégicos de la organización.

Se han seleccionado cuatro criterios de evaluación de calidad de proyectos informáticos y un criterio que se alinea a los planes estratégicos, los criterios son los siguientes:

- **Fiabilidad:** La capacidad de la implementación para mantener un nivel específico de funcionamiento cuando se está utilizando bajo condiciones especificadas.
- **Eficiencia:** La capacidad de la implementación para proveer un desempeño adecuado, de acuerdo a la cantidad de recursos utilizados y bajo las condiciones planteadas.
- **Eficacia:** La capacidad de la implementación para permitir a los usuarios lograr las metas especificadas con exactitud e integridad, en un contexto especificado de uso.
- **Usabilidad:** La capacidad de la implementación de ser entendido, aprendido, usado y atractivo al usuario, cuando es utilizado bajo las condiciones especificadas.
- **Adaptabilidad para nuevos servicios:** La capacidad de la solución implementada para lograr ser adaptada y poder solucionar problemas diversos, buscando generar nuevos servicios para la organización.

4.2. INFORMACIÓN DE SITUACIÓN ACTUAL

Como comentario inicial, cabe indicar que **gestionona** al iniciar el desarrollo del proyecto simultáneamente empezó la negociación del contrato marco para el pago del servicio de control de fraude para detección de hurto de línea, antes el cliente pagaba por cantidad de FTEs (Full Time Equivalent - Equivalente a Tiempo Completo), con cual es indiferente a la efectividad de las fuentes generadoras de alertas, el nuevo driver de pago sería por caso de hurto de línea detectado.

Empezaremos diagramando la situación anterior a la implementación del proyecto del proceso de Control de Fraude, donde se indican las fuentes de datos y los principales procesos además de los flujos de información más importantes, SAMARA es el sistema antifraude InHouse desarrollado por el equipo de Lucha contra el Fraude, brinda el soporte a todo el proceso de control de fraude, como se observa se contaba con muchas fuentes, donde el principal problema es que las alertas generadas no tenían la efectividad superior al 20%.

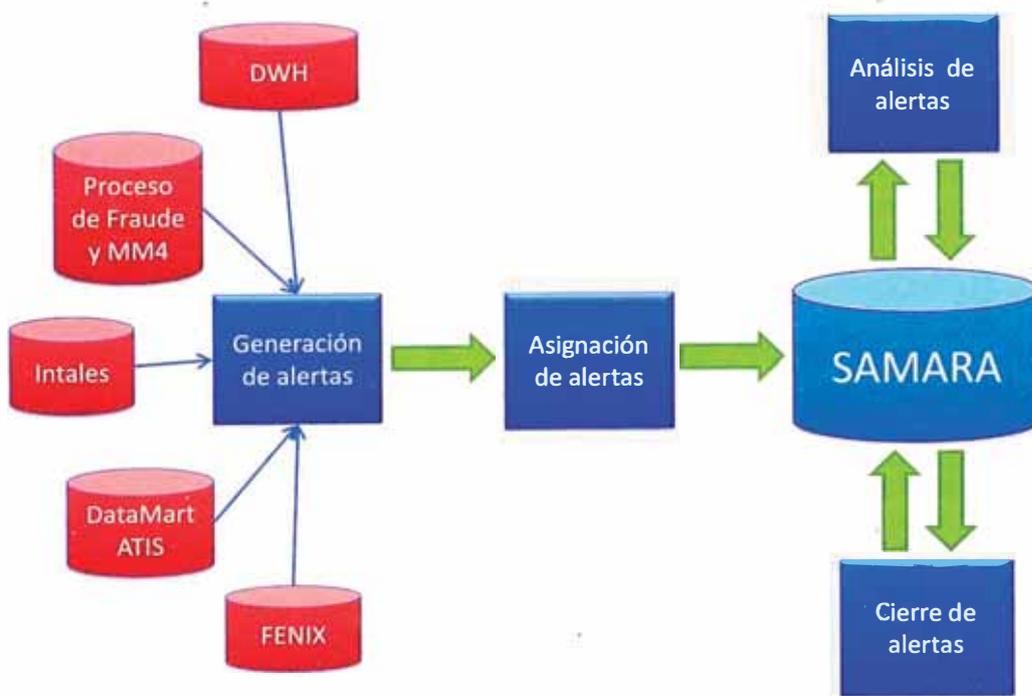


Gráfico 4.1: Diagrama de la Situación anterior a la implementación del Proyecto (Fuente: LCF, Elaboración: Propia).

Continuaremos realizando el análisis económico para nuestro proyecto. Se necesita realizar el flujo de caja, para lo cual se realiza el cálculo de los costos iniciales del proyecto, se cuenta con los siguientes:

- **Capacitación:** Se considera toda la inversión inicial realizada para que el equipo del proyecto tenga los conocimientos necesarios para empezar. Se llevo el Curso básico y avanzado de Análisis predictivo – Minería de datos.
- **Hardware:** Se considera toda la infraestructura técnica adquirido para el desarrollo e implementación del proyecto. Se adquirió Tres PCs HP con procesadores Core2Duo de 2.1 GHz, 4 GB de memoria RAM, 320 GB HD y una PC HP con procesador QuadCore de 2.0 GHz, 4 GB de memoria RAM, 500 GB HD, etc.
- **Software:** Se considera las licencias del software adquirido para el desarrollo e implementación del proyecto. Se adquirió la licencia del SQL Server, MS Office 2007, Windows XP, etc.
- **Personal:** Se considera el pago mensual de un Jefe de Proyecto, un analista, un estadístico y dos desarrolladores, por siete meses que fue la duración del proyecto.

Para el cálculo de los beneficios, se va considerar el nuevo contrato marco con el nuevo driver de pago:

- **Cantidad de casos detectados de hurto de línea:** Se considero para el cálculo el promedio de casos detectados por hurto de línea en el periodo de abril 2011 a abril 2012, por el costo establecido en el nuevo contrato marco por cada caso de hurto de línea detectado.

A continuación se describe el flujo de caja final, donde se ha calculado el VAN y el TIR, además de evaluarlos según los resultados obtenidos:

DESCRIPCION	Año 0	Año 1	Año 2	Año 3	Año 4
INGRESOS	0	352,800	352,800	352,800	352,800
Por casos detectados	0	352,800	352,800	352,800	352,800
Por Ahorros		18,000	18,000	18,000	18,000
GASTOS	195,314	181,801	181,801	181,801	181,801
Hardware	9,213	0	0	0	0
Software	5,101	5,101	5,101	5,101	5,101
Personal	171,000	175,200	175,200	175,200	175,200
Mantenimiento		1,500	1,500	1,500	1,500
Capacitación	10,000	0	0	0	0
UTILIDAD BRUTA	-195,314	170,999	170,999	170,999	170,999
GASTOS OPERATIVOS	11,000	11,000	11,000	11,000	11,000
EBITDA	-206,314	159,999	159,999	159,999	159,999
DEPRECIACION		2,400	2,400	2,400	2,400
UTILIDAD OPERATIVA NETA	-206,314	157,599	157,599	157,599	157,599
IMPUESTOS	61,894	47,280	47,280	47,280	47,280
UTILIDAD OPERATIVA DESPUES DE IMPUESTOS	-268,208	110,319	110,319	110,319	110,319
FLUJO DE CAJA ACUMULADO	-268,208	-157,889	-47,570	62,750	173,069
Principales Indicadores					
	TIR: 23%		VAN: S/. 308,861		

Cuadro 4.1: Flujo de Caja del Proyecto (Fuente y Elaboración: LCF).

Como se puede observar se obtuvo el VAN de S/. 308,861 y TIR de 23%, con lo cual podemos concluir:

- Dado que el VAN > 0 se acepta invertir en este proyecto, éste importe representa la ganancia adicional o extraordinaria que recibirá el inversionista, dada por el rendimiento de la inversión.
- Si TIR = 23% ≥ 15% (tasa de descuento para el VAN), es decir se acepta el proyecto, debido a que se confirma la rentabilidad.

En resumen, se recomienda que se realice la implementación y desarrollo del proyecto por la evaluación económica.

Ahora realizaremos una evaluación de los criterios detallados en el punto anterior mediante el siguiente cuadro que compara la situación antes y después de la implementación del proyecto:

Criterio de evaluación	Antes de la Implementación	Después de la Implementación
Fiabilidad	El sistema presentaba caídas durante el procesamiento excesivo de las diversas fuentes de datos que servía para la generación de las alertas. Cuando existía simultaneidad de los reportadores con los analistas de fraude, el sistema bajaba su performance.	Se redujo las fuentes de datos necesarias, con lo cual se logro optimizar la actividad de carga de data, además de la generación de alertas redujo el tiempo de ejecución total.
Eficiencia	Los tiempos de carga de datos en los inicios de cada semana eran considerablemente largos, una vez cargado los datos la generación y el resto de procesos, manejaban tiempos de respuesta óptimos. El sistema mostraba la información necesaria en una misma pantalla, con lo que los analistas de fraude reducían sus tiempos de análisis por no estar navegando por diversos sistemas extrayendo o validando dicha información.	Maneja tiempos de respuesta óptimos, en los diversos procesos del sistema. Se realizo la distribución de la base de datos para evitar que los reportadores bajen la performance del sistema y reducir el tiempo de ejecución de los reportes y consultas diversas.
Eficacia	Usuarios necesitaban procesar información por fuera del sistema, para luego cargarla y que sirva de input al proceso de generación de alertas.	Se logro automatizar y depurar la carga de datos, evitando procesamiento manual fuera del sistema.
Usabilidad	Se necesitaba de programar y configurar las cargas de las diversas fuentes de datos, sin embargo para nuevas fuentes se necesitaba contar con un conocimiento especializado para realizarlo. Para el analista de fraude el sistema si es intuitivo y sencillo.	Con el modelo de minería de datos y su despliegue embebido dentro del sistema, los asignadores solo deben de correr la carga de datos y crear la programación de asignación, para poder generar sin mayores complicaciones la generación de alertas.
Adaptabilidad para nuevos servicios	El sistema ha sido estructurado para soportar el control de fraude en telecomunicaciones, para poder adaptarlo a nuevos servicios se hace muy rigido o demasiado laborioso. Se utilizaban perfiles y patrones que los expertos desarrollaban para problemas particulares de fraude en telecomunicaciones.	La experiencia adquirida brinda la oportunidad a tggestiona de brindar servicios de consultoría, desarrollo e implementación de proyectos de minería de datos, en diversos sectores o rubros comerciales, donde el entregable final sería un modelo óptimo que logre solucionar el problema específico con que cuenta los clientes.

Cuadro 4.2: Criterios de evaluación para antes y después de la implementación del proyecto (Fuente y Elaboración: Propia).

4.3. RESULTADOS DE LA SOLUCIÓN PLANTEADA

La implementación del proyecto logró conseguir muchas mejoras en la generación de alertas que es el principal proceso para incrementar la efectividad en la detección del hurto de línea, a continuación se diagrama la situación posterior a la implementación del proyecto del proceso de Control de Fraude:

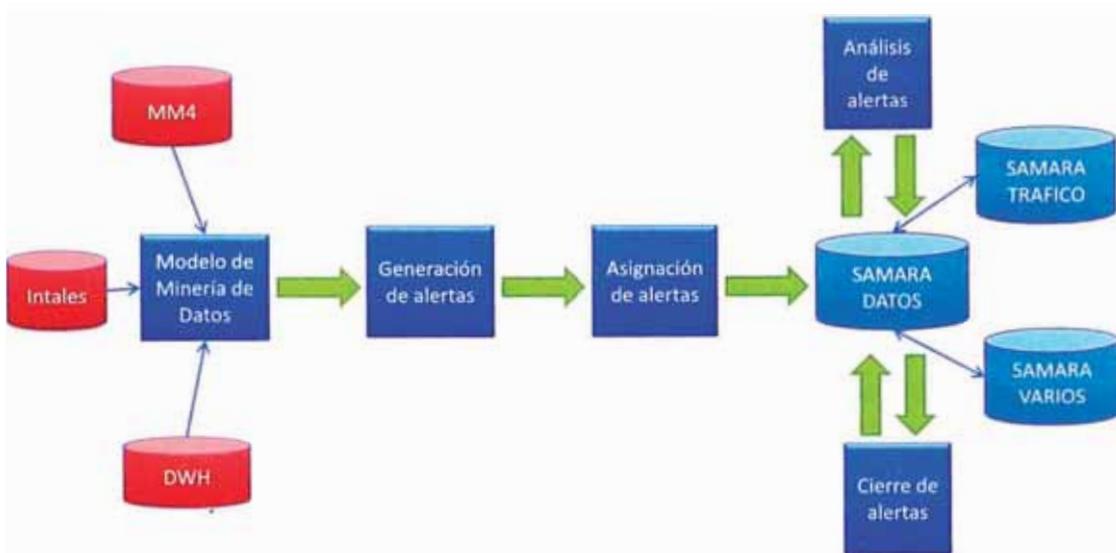


Gráfico 4.2: Diagrama de la Situación posterior a la implementación del Proyecto (Fuente: LCF, Elaboración: Propia).

Entre las mejoras logradas posterior a la implementación del proyecto, se pueden resaltar las siguientes:



Gráfico 4.3: Principales beneficios posterior a la implementación del proyecto (Fuente: LCF, Elaboración: Propia).

- Se incremento la efectividad de las alertas de un 20% a un 70% para detectar hurto de línea, lo cual implica mayores ingresos con el nuevo contrato marco, donde se cobra al cliente por caso detectado.
- La técnica de minería de datos solo se implementa en el proceso de generación de alertas, sin embargo las mejoras impactan en todo el proceso de Control de Fraude.
- La nueva estructura de trabajo logra incrementar el performance del Sistema, principalmente en el análisis y registro de alertas y en la generación de reportes varios.

- Al tener éxito en la implementación del modelo de minería de datos para la detección de hurto de línea, se empezó con el diseño de nuevos modelos para otras tipologías de fraude que cuentan con una baja efectividad en la detección, tales como arbitraje, ilegal landing, hackeo de centralitas y fraude por suscripción con consumo anómalo.
- Al contar con la experiencia de la implementación de un proyecto de minería de datos, se ha observado un gran potencial para poder desarrollar nuevos proyectos en otros campos como la Banca, Retails, etc., el cual brinda la oportunidad de ofrecer el nuevo servicio “Diseño y detección de Patrones bajo la técnica de minería de datos” al mercado.
- Los beneficios que se consiguieron en el servicio de Control de fraude para detección de hurto de línea, indirectamente le trajo beneficios al cliente como reducción de reclamos por desconocimiento de llamadas, reducción de la pérdida promedio por caso detectado, reducción de las inspecciones técnicas que terminaban siendo falsos positivos, etc.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES:

- Las organizaciones en la actualidad se han vuelto dependientes de los sistemas de información y las tecnologías de información, principalmente porque a través de su uso logran la automatización de procesos operativos, las herramientas de información necesaria para la toma de decisiones y lograr contar con nuevas ventajas competitivas. La minería de datos nos permite capitalizar la información, generando conocimiento para tomar decisiones más inteligentes.
- El gran volumen de información que presenta el sector de las telecomunicaciones demanda la utilización de técnicas y/o herramientas altamente eficientes y escalables.
- Se ha demostrado que el uso de la técnica de minería de datos puede ayudar considerablemente a determinar las causas de los problemas identificados y a comprender con más profundidad el proceso de generación de alertas, además que ha demostrado ser viable y posible de implementar, sin embargo no se puede aplicar cualquier algoritmo a ciegas o forzar encontrar relaciones donde no existen.

- Se propuso el diseño de un modelo de minería de datos utilizando redes neuronales para la detección de fraude de hurto de línea basada en la hipótesis que estos casos presentan un comportamiento particular, que muchas veces escapa de la simple visualización.
- Los resultados, han sido excelentes, aunque pueden ser aún mejorados si se utilizan bases de datos con data histórico mayor a un año, además de probar los correctos tipos y formas de entrenamiento de los modelos de minerías de datos.
- El algoritmo de las redes neuronales son uno de los enfoques más utilizados para el aprendizaje inductivo y han demostrado un buen comportamiento predictivo en una gran variedad de problemas en diversas industrias. Sin embargo cuenta con una gran limitación, los resultados de los modelos suelen ser algo incomprensibles, porque el proceso de aprendizaje de las redes neuronales al momento de extraer las reglas de clasificación no se puede representar todo el modelo embebido en la red neuronal.
- Las efectividades de los modelos de minería de datos fueron muy superiores comparándolos con los criterios tradicionales que se utilizaban para detectar hurto de línea.

RECOMENDACIONES:

- Para poder aplicar la técnica de minería de datos, se necesitan tener algunas consideraciones básicas como contar con datos organizados, limpios y procesados porque la calidad de la información determinara la calidad del resultado,

- Para obtener los resultados esperados se debe de contar con objetivos específicos bien definidos, además tener en cuenta que no es una tecnología mágica.
- Para poder utilizar la técnica de minería de datos, es necesario hacer un trabajo involucrando a los expertos del negocio. En la construcción de los modelos, se necesita generar un conjunto de datos para el entrenamiento, validación y pruebas con variables significativas que puedan entregar un resultado que nos sirva para solucionar nuestro problema planteado.
- Se recomendó realizar una renegociación del contrato marco para el pago del servicio de control de fraude para detección del hurto de línea, de pago por cantidad de FTEs (Full Time Equivalent) a un pago por caso de hurto de línea detectado.
- Se recomendó implementar la técnica de minería de datos para incrementar la efectividad en la detección de otras tipologías de fraude como arbitraje, ilegal landing, hackeo de centralitas y fraude por suscripción con consumo anómalo.
- Debido al gran potencial que presenta la técnica de minería de datos se recomendó desarrollar proyectos en otros campos como la Banca, Retails, Manufactura, etc.

GLOSARIO DE TERMINOS

- **Algoritmos de minería de datos:** El algoritmo de minería de datos es el mecanismo que crea modelos de minería de datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos, buscando patrones y tendencias específicos. Después, el algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de minería de datos.
- **Arbitraje:** Uso indebido de los servicios telefónicos realizando llamadas entre dos países utilizando un tercero de puente, buscando una ganancia aprovechando la diferencia de precios de las llamadas entre el país origen al país puente y el país puente al país destino, siendo estas llamadas mas económicas que la llamada directa del país origen al país destino.
- **Averías:** Daño o desperfecto que presenta un servicio telefónico debido a diversos factores, desde equipos dañados, obsoletos o hasta daños en las facilidades técnicas (planta interna o externa).
- **Back Office:** Es la parte de las empresas de telecomunicaciones donde tienen lugar las tareas destinadas a gestionar las ventas y con las cuales el cliente no necesita contacto directo. Se encarga de registrar y validar las ventas, los reclamos, etc.

- **Central PBX:** Es un dispositivo que actúa como una ramificación de la red primaria pública de teléfonos, por lo que los usuarios no se comunican directamente al exterior mediante líneas telefónicas convencionales, sino que al estar el PBX directamente conectado a la RTC (red telefónica pública), será esta misma la que enrute la llamada hasta su destino final mediante enlaces unificados de transporte de voz llamados líneas troncales.
- **Conocimiento:** Hechos o datos de información adquiridos por una persona a través de la experiencia o la educación, la comprensión teórica o práctica de un tema de la realidad, incluye el "saber qué" (know what), el "saber cómo" (know how) y el "saber dónde" (know where).
- **Consumo anómalo:** Es aquel consumo que se encuentra muy por encima del consumo promedio normal, en el caso de la telefonía fija es aquel servicio telefónico que presenta elevada cantidad de llamadas o minutos.
- **Control de fraude:** Acciones que realizan las empresas de telecomunicaciones para mitigar las actividades que realizan los fraudulentos que buscan usufructuar con sus productos y servicios.
- **CRISP-DM (Cross Industry Standard Process for Data Mining – Cross Proceso Estándar de la Industria para Minería de datos):** Es una de las metodología que se utilizan para desarrollar la técnica de minería de datos, está descrita en términos de un modelo de proceso jerárquico, consistente en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico): fase, tarea genérica, tarea especializada, e instancia de procesos.

- **DataWareHouse:** Es un repositorio de datos de muy fácil acceso, alimentado de numerosas fuentes, transformadas en grupos de información sobre temas específicos de negocios, para permitir nuevas consultas, análisis, reporteador y decisiones.
- **DISA (Direct Inward System Access - Acceso directo entrante al sistema):** Es una prestación de las Centralita Telefónicas convencionales, que ofrece la posibilidad de acceder a una extensión determinada conectada a la Centralita desde el exterior sin necesidad de pasar por la operadora o telefonista.
- **Fraude por Suscripción:** Tipo de fraude que se presenta cuando se utiliza documentación falsa o de terceros con la finalidad de que los cargos y la facturación se registren a nombre de otra persona. Por lo general, cuando se adquieren servicios de telefonía, el defraudador consume lo máximo posible y posteriormente se retira del local arrendado.
- **Grupo Telefónica:** Es un grupo de empresas que se desarrolla en el rubro de servicios de telecomunicaciones (telefonía fija, telefonía móvil, ADSL, FTTH, etc.) multinacional con sede central en Madrid, España, es la quinta compañía de telecomunicaciones en tamaño e importancia en el mundo.
- **Hacker:** Es una persona que pertenece a una de estas comunidades de gente apasionada por la seguridad informática, esto concierne principalmente a entradas remotas no autorizadas por medio de redes de comunicación ("Black hats"). Pero también incluye a aquellos que depuran y arreglan errores en los sistemas ("White hats") y a los de moral ambigua como son los "Grey hats".

- **Hurto de Línea:** Este un tipo de fraude que puede ser interno o externo, se da cuando servicios telefónicos activos con asignación a usuarios son cambiadas de facilidades técnicas sin autorización del suscriptor o de la empresa local proveedora del servicio.
- **Ilegal landing:** Uso indebido que encamina directamente el tráfico que viene del exterior hacia las centrales locales, sin pasar por la central de tráfico internacional, es decir, se evita la tarifación de la llamada internacional, y se la convierte en una llamada local.
- **KDD (Descubrimiento de Conocimiento en Bases de Datos):** Es el Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos. El objetivo fundamental del KDD es encontrar conocimiento útil, válido, relevante y nuevo sobre un fenómeno o actividad mediante algoritmos eficientes, dadas las crecientes órdenes de magnitud en los datos.
- **Minería de datos:** Consiste en la extracción no trivial de información que reside de manera implícita en los datos. Dicha información era previamente desconocida y podrá resultar útil para algún proceso. En otras palabras, la minería de datos prepara, sondea y explora los datos para sacar la información oculta en ellos.
- **SEMMA (Sample, Explore, Modify, Model, Assess - Muestra, Estudiar, Modificar, Modelar, Evaluar):** Metodología que se refiere al proceso central de la realización de minería de datos. A partir de una muestra estadísticamente representativa de sus datos, SEMMA hace que sea fácil de aplicar técnicas estadísticas de exploración y visualización, seleccionar y transformar las variables predictivas más importantes, el

modelo de las variables para predecir los resultados y confirmar la exactitud de un modelo.

- **Telecomunicaciones:** El término que cubre todas las formas de comunicación a distancia, incluyendo radio, telegrafía, televisión, telefonía transmisión de datos e interconexión de computadoras a nivel de enlace. Según la Unión Internacional de Telecomunicaciones, las telecomunicaciones son "toda transmisión, emisión o recepción de signos, señales, datos, imágenes, voz, sonidos o información de cualquier naturaleza que se efectúa a través de cables, medios ópticos, físicos u otros sistemas electromagnéticos".
- **Teléfono Público:** Es un teléfono que funciona introduciendo monedas, tarjetas telefónicas o tarjetas de crédito. Normalmente el precio de la llamada es más caro que el fijo de casa pero más barato del teléfono móvil. Normalmente se encuentran ubicados en lugares públicos como intercambiadores de transporte, plazas o centros comerciales.

BIBLIOGRAFIA

- [TGES01]: TGESTIONA, Historia, Lima - Perú, 2011, Disponible en la Web:
<<http://moss.tgestiona.com.pe/nosotros/historia/paginas/default.aspx>>.
- [TGES02]: TGESTIONA, Soluciones, Lima - Perú, 2011, Disponible en la Web: <<http://moss.tgestiona.com.pe/soluciones/paginas/default.aspx>>.
- [TGES03]: TGESTIONA, Experiencia - Clientes, Lima - Perú, 2011, Disponible en la Web:
<<http://moss.tgestiona.com.pe/experiencia/paginas/clientes-telecomunicaciones.aspx>>.
- [TGES04]: TGESTIONA, Misión y visión, Lima - Perú, 2011, Disponible en la Web:
<<http://moss.tgestiona.com.pe/nosotros/misionvision/paginas/default.aspx>>
- [DEPE01]: GENTEUCV.CC.CC, Historia de las Telecomunicaciones en el Perú, Lima Perú, 2008, Disponible en la Web:
<<http://comunicacionesiia.blogspot.com/2008/12/historia-de-las-telecomunicaciones-en.html>>.
- [SUPE01]: ING. MARÍA JOSÉ MEZA AYALA, Investigación Especial en Telecomunicaciones, Profesional Técnico - Dirección General de Investigación en Telecomunicaciones, Pág. 13, Quito - Ecuador, 2008.
- [SUPE02]: ING. MARÍA JOSÉ MEZA AYALA, Investigación Especial en Telecomunicaciones, Profesional Técnico - Dirección General de Investigación en Telecomunicaciones, Pág. 25, Quito - Ecuador, 2008.

- [SUPE03]: ING. MARÍA JOSÉ MEZA AYALA, Investigación Especial en Telecomunicaciones, Profesional Técnico - Dirección General de Investigación en Telecomunicaciones, Pág. 20, Quito - Ecuador, 2008.
- [SCRI01]: PEDRO PIZARRO ALMEIDA, Tipos de Fraude en Telecomunicaciones por fraude, Pág. 06, Guayaquil - Ecuador, 2011.
- [SCRI02]: PEDRO PIZARRO ALMEIDA, Tipos de Fraude en Telecomunicaciones por fraude, Pág. 09, Guayaquil - Ecuador, 2011.
- [SCRI03]: PEDRO PIZARRO ALMEIDA, Tipos de Fraude en Telecomunicaciones por fraude, Pág. 11, Guayaquil - Ecuador, 2011.
- [DOFY01]: DORIAN PYLE, Data Preparation for Data Mining, Pág. 17, San Francisco - USA, 1999.
- [KDNU01]: KDNUGGETS, Data Mining Methodology, Boston - USA, 2007, Disponible en la Web: <http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm>.
- [SPSS01]: SPSS, CRISP-DM 1.0, USA, 2000, Disponible en la Web: <<ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>>.
- [SASS01]: SAS, SEMMA, London - United Kingdom, 2011, Disponible en la Web: <<http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html>>.
- [KDNU02]: KDNUGGETS, Industries where you applied Analytics / Data Mining in 2011, Boston - USA, 2011, Disponible en la Web: <<http://www.kdnuggets.com/polls/2011/industries-applied-anaytics-data-mining.html>>.

ANEXOS

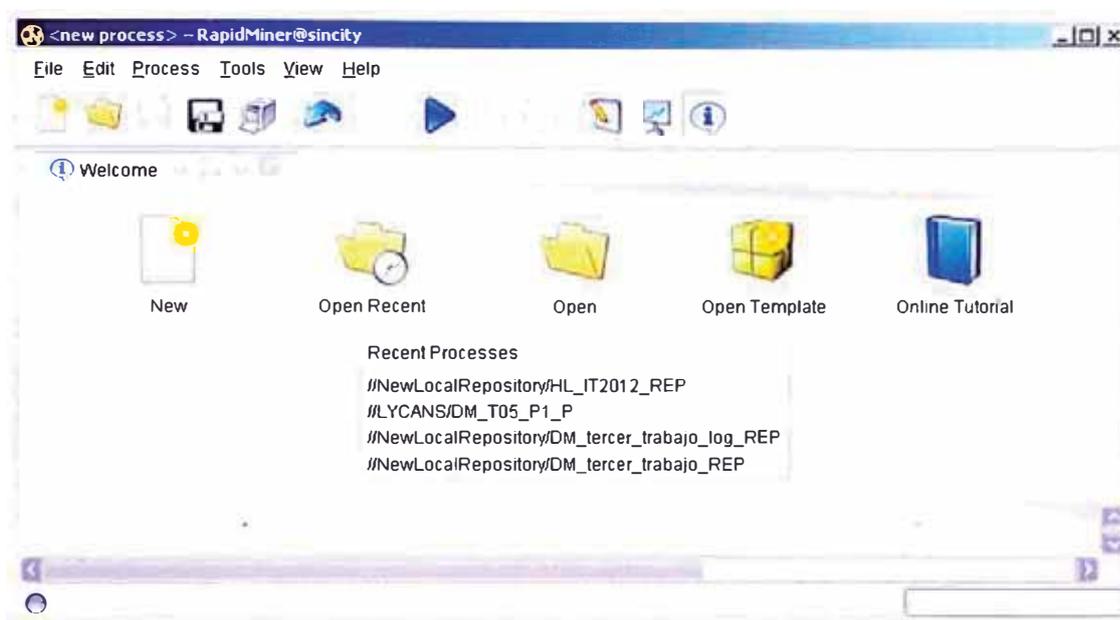


Gráfico A.1: Pantalla de inicio del RapidMiner (Fuente: RapidMiner, Elaboración: Propia).

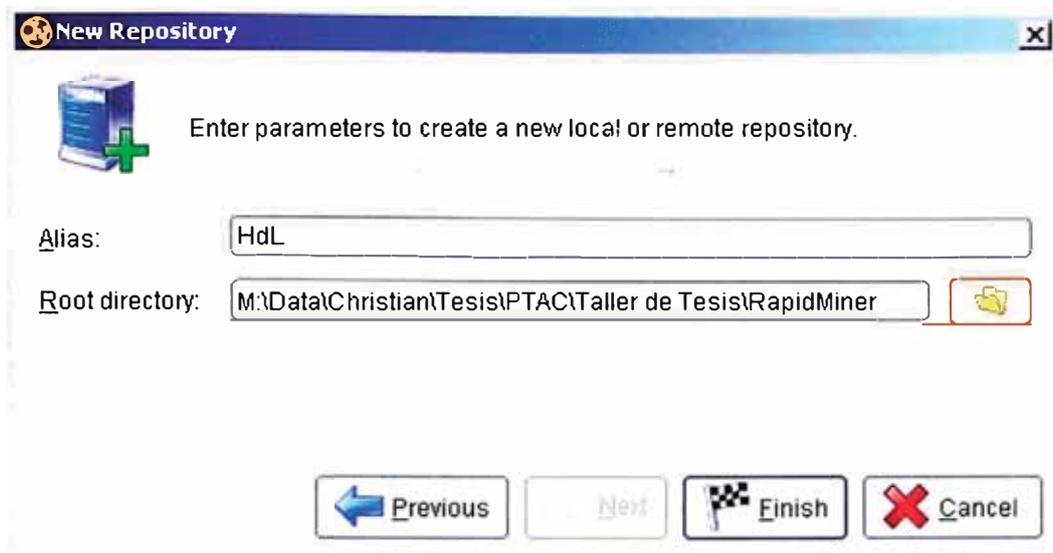


Gráfico A.2: Creación de un Repositorio para los datos de Hurto de Línea (Fuente: RapidMiner, Elaboración: Propia).

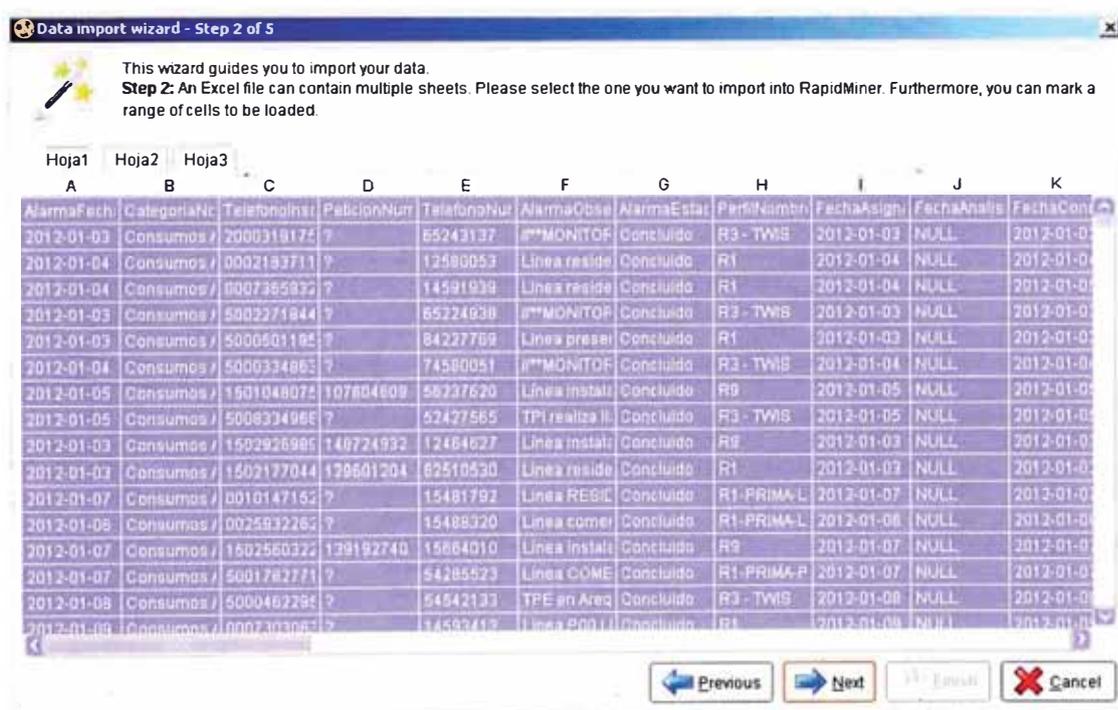


Gráfico A.3: Carga de datos de una de las fuentes (Fuente: RapidMiner, Elaboración: Propia).

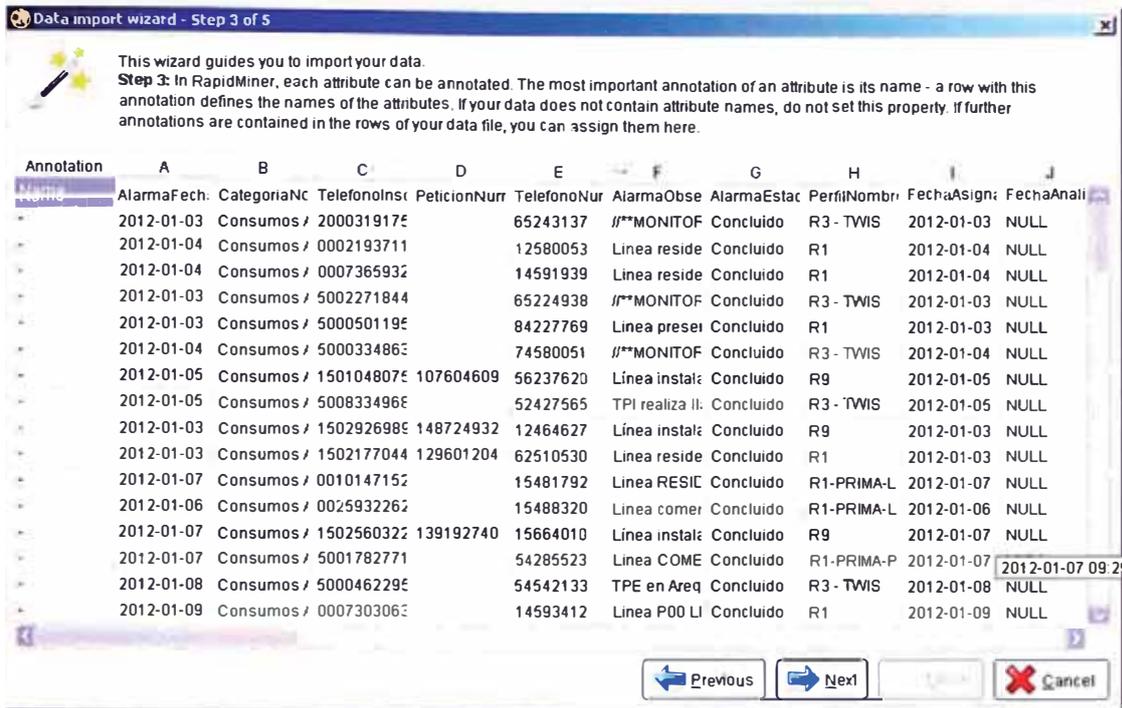


Gráfico A.4: Selección de la cabecera y de los registros de los datos de la fuente (Fuente: RapidMiner, Elaboración: Propia).

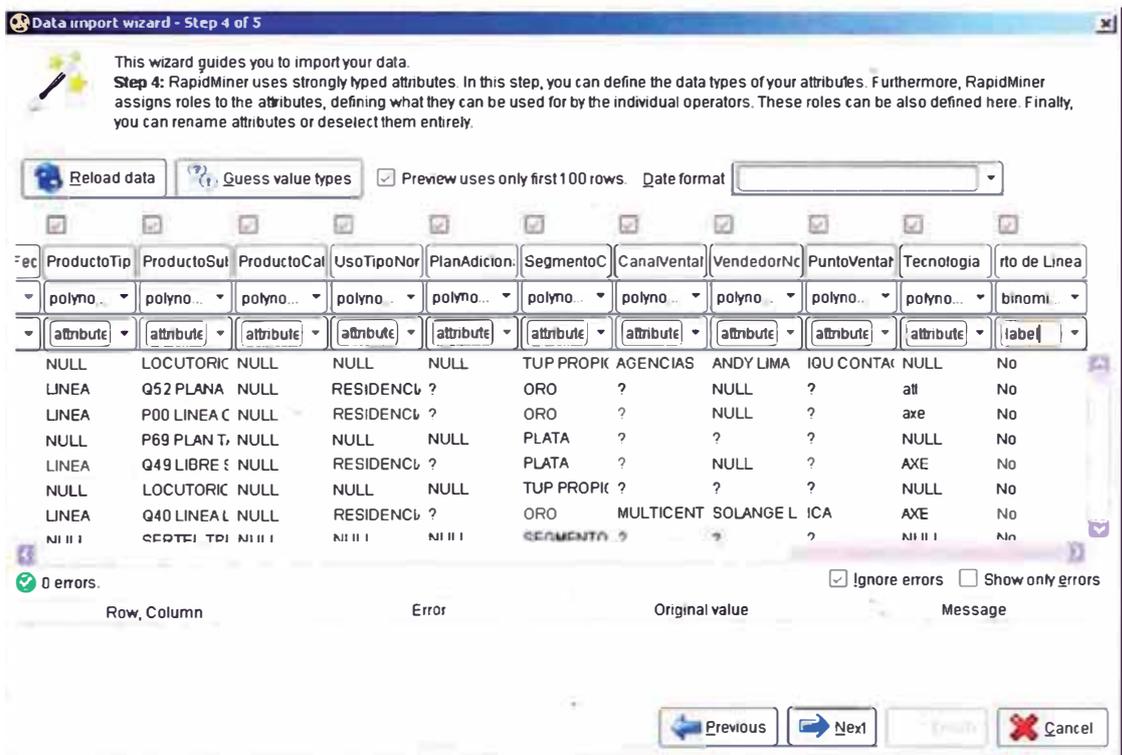


Gráfico A.5: Selección de los tipos de datos de los registros de los datos de la fuente (Fuente: RapidMiner, Elaboración: Propia).

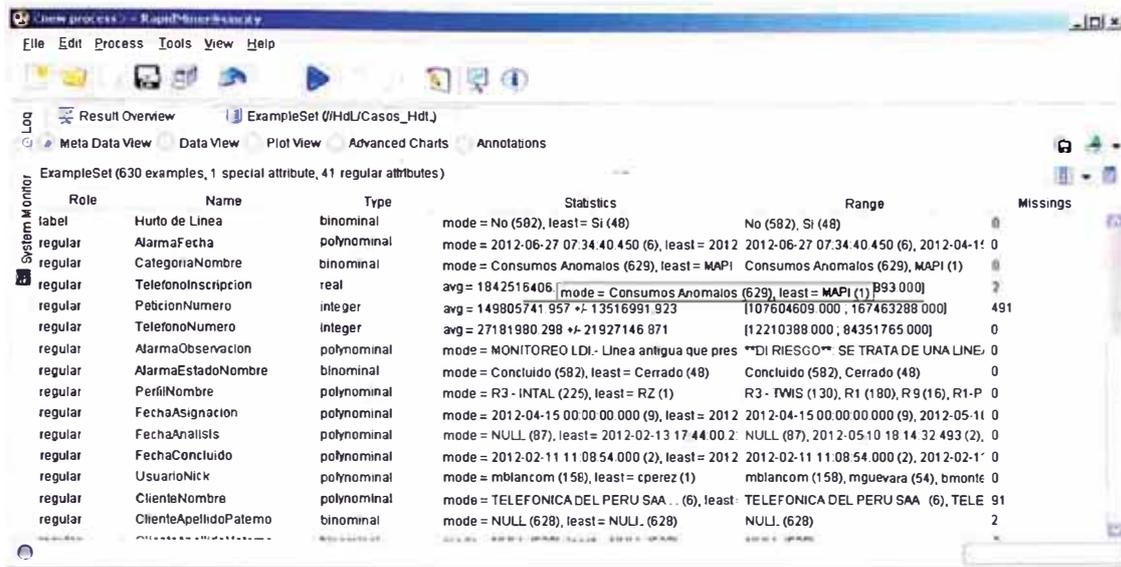


Gráfico A.6: Estadística descriptiva de los datos de la fuente (Fuente: RapidMiner, Elaboración: Propia).

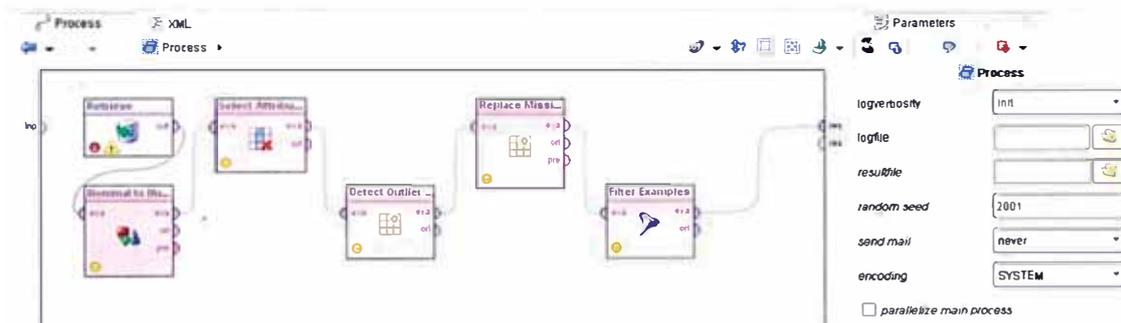


Gráfico A.7: Calidad, limpieza y transformación de los datos (Fuente: RapidMiner, Elaboración: Propia).

ExampleSet (630 examples, 2 special attributes, 41 regular attributes)

Role	Name	Type	Statistics	Range	Missings
label	Hurto de Linea	binominal	mode = No (582), least = Si (48)	No (582), Si (48)	0
outlier	outlier	binominal	mode = false (620), least = true (10)	false (620), true (10)	0
regular	AlarmaFecha	numeric	avg = 261.465 +/- 146.540	[0.000 ; 516.000]	0
regular	CategoriaNombre	numeric	avg = 0.002 +/- 0.040	[0.000 ; 1.000]	0
regular	AlarmaObservacion	numeric	avg = 309.178 +/- 179.215	[0.000 ; 620.000]	0
regular	AlarmaEstadoNombre	numeric	avg = 0.076 +/- 0.266	[0.000 ; 1.000]	0
regular	PerfilNombre	numeric	avg = 4.717 +/- 4.456	[0.000 ; 12.000]	0
regular	FechaAsignacion	numeric	avg = 249.297 +/- 140.914	[0.000 ; 501.000]	0
regular	FechaAnalisis	numeric	avg = 233.043 +/- 171.993	[0.000 ; 540.000]	0
regular	FechaConcluido	numeric	avg = 309.683 +/- 179.870	[0.000 ; 622.000]	0
regular	UsuarioNick	numeric	avg = 4.737 +/- 4.365	[0.000 ; 14.000]	0
regular	ClienteNombre	numeric	avg = 256.699 +/- 137.813	[0.000 ; 519.000]	0
regular	ClienteApellidoPaterno	numeric	avg = 0 +/- 0	[0.000 ; 0.000]	0
regular	ClienteApellidoMaterno	numeric	avg = 0 +/- 0	[0.000 ; 0.000]	0
regular	ClienteDocTipo	numeric	avg = 0.286 +/- 0.448	[0.000 ; 1.000]	0
regular	ClienteDireccion	numeric	avg = 305.395 +/- 178.871	[0.000 ; 616.000]	0
regular	ClienteDistrito	numeric	avg = 43.022 +/- 33.026	[0.000 ; 145.000]	0
regular	ClienteProvincia	numeric	avg = 9.318 +/- 15.611	[0.000 ; 72.000]	0
regular	ClienteDepartamento	numeric	avg = 5.278 +/- 8.400	[0.000 ; 44.000]	0
regular	ClienteCodigoPostal	numeric	avg = 55.217 +/- 37.428	[0.000 ; 158.000]	0
regular	LimaProvincia	numeric	avg = 0 +/- 0	[0.000 ; 0.000]	0

Gráfico A.8: Datos después de la limpieza y transformación (Fuente: RapidMiner, Elaboración: Propia).

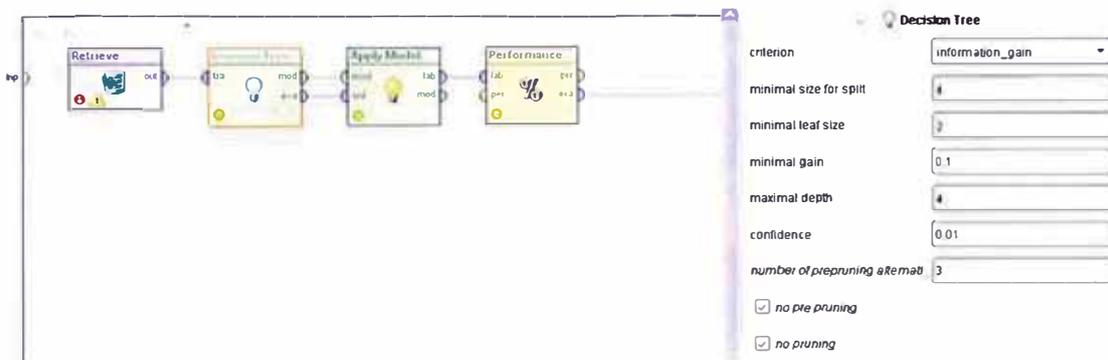


Gráfico A.9: Aplicación del modelo de Árbol de Decisiones (Fuente: RapidMiner, Elaboración: Propia).

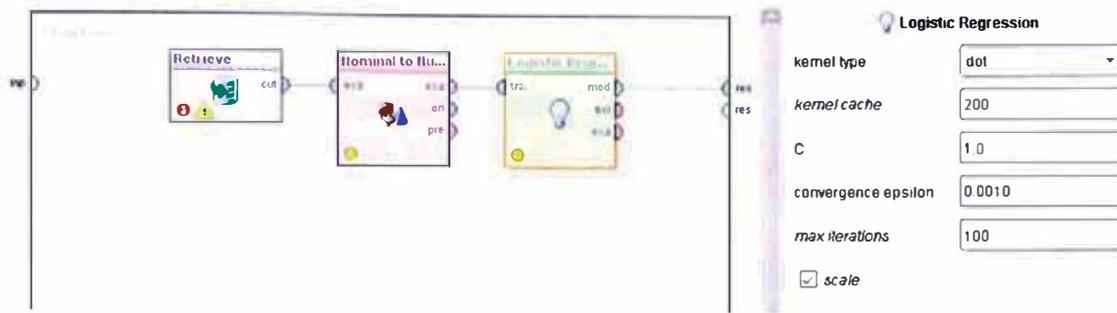


Gráfico A.10: Aplicación del modelo de Regresión Logística (Fuente: RapidMiner, Elaboración: Propia).

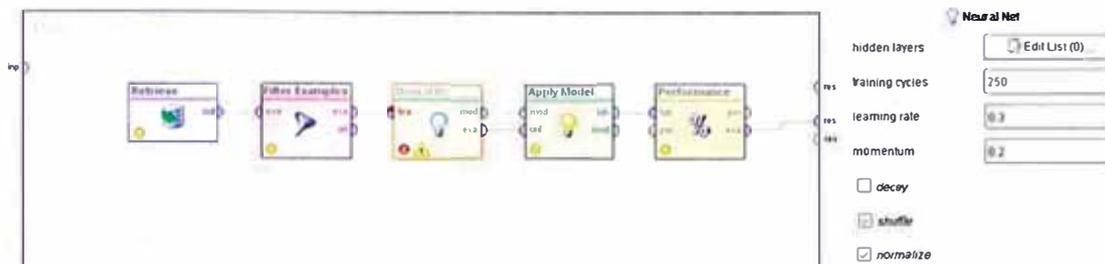


Gráfico A.11: Aplicación del modelo de Redes Neuronales (Fuente: RapidMiner, Elaboración: Propia).

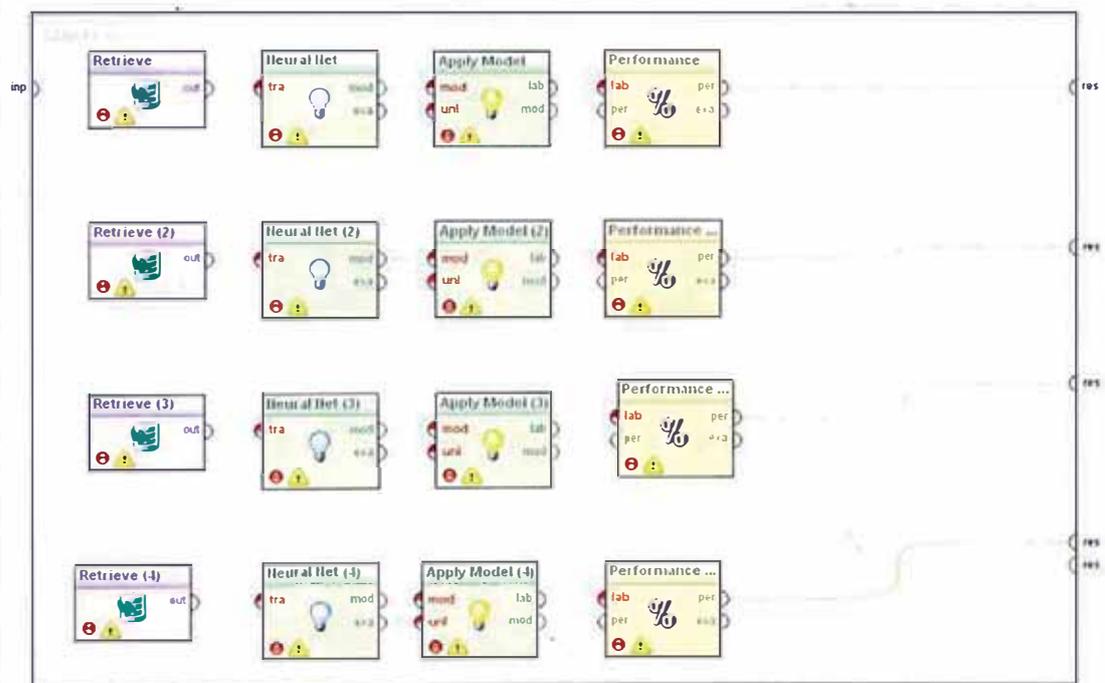


Gráfico A.11: Optimización del modelo de Redes Neuronales (Fuente: RapidMiner, Elaboración: Propia).