



UNIVERSIDAD
AUTÓNOMA
DE ICA

FACULTAD DE INGENIERIA, CIENCIAS Y ADMINISTRACIÓN

TESIS

DETECCIÓN DEL PROBLEMA DE FRAUDE EN TELEFONÍA
CELULAR, CON EL USO DE REDES NEURONALES EN LIMA
DISTRITO, AÑO 2018.

TRABAJO REALIZADO PARA OBTAR EL TÍTULO PROFESIONAL
DE: INGENIERO DE SISTEMAS

PRESENTADO POR:

José Luis Aguilar Montesinos

ASESOR: Roberto Buendia Aparcana

CHINCHA- ICA -PERÚ

2019

DEDICATORIA:

Para mis padres, hermanos y mi toda
familia por el apoyo incondicional
que siempre me han dado,
que me han permitido
lograr mis
objetivos.

AGRADECIMIENTOS

A mi familia, mis amigos que compartimos carpeta, a
mis profesores y al personal administrativo,
quienes en todo momento asumimos
problemas y responsabilidades,
todos tenemos un común objetivo
que es convertirnos en un
profesional en sistemas.
A todos ellos siempre
los recordaré.

RESUMEN:

El objetivo que se busca con este trabajo de investigación, tiene como objetivo fundamental I detectar en forma técnica, cómo las variaciones en el consumo en el uso de telefonía celular fuera de lo normal, considerándose este punto como nuestro objetivo más importante. Se observa la filtración de este servicio por actos ilegales de los inescrupulosos usuarios, en perjuicio de los propietarios de líneas; es por esta razón que a través de esta investigación se pretende detectar las técnicas de la filtración para recomendar las alternativas de solución.

En el desarrollo de este trabajo el investigador pone de manifiesto, cómo se construyen estructuras de información que presentan datos recientes e históricos de cada uno de los usuarios, en donde hacemos uso de información que contiene una llamada, se observa en la determinación de las llamadas, decir de si las llamadas LOC, NAT e INT., resultando complejo también compleja la construcción de una función con tantas variables de entradas paramétricas, que no siempre son conocidas.

Como se ha determinado la presencia de diferentes métodos de detección del fraude, todas ellas tienen resultados eficientes, arrojando siempre en cantidades incompletas a las que los autores lo conocen como picos trabajan con picos de consumo o reglas fijas, para lo cual la metodología aplicada tiene que estar de acuerdo a los requerimientos; por lo mismo que es preciso mencionar que se trata de una investigación básica y tal como sostenía Hernández, R. (2010,Pág. 192), Las investigaciones básicas se caracterizan por que son dirigidas especialmente distinguir y aplicar como las leyes generales de los fenómenos estudiados, aspecto q que nos permiten determinar las teorías más significantes y discretas.

La solución que presentamos han sido elaborados haciendo uso de la tecnología de redes neuronales no supervisadas, en particular las redes SOM. Entre ellas se presenta:

- Los gráficos en donde se observa los patrones de los listados creados con sus respectivos valores.

- Las pruebas de las llamadas Locales, Nacionales e Internacionales.

Del mismo modo se concluye en el siguiente:

- a) La estructura que deben tener los patrones de llamadas LOC, NAT, e INT.
- b) Se ha establecido que los registros de estructura para el servicio LOC debe ser de 144 patrones: NAT debe ser 64 patrones y del INT, 64 patrones.

SUMMARY

The objective that is sought with this research work, has as its fundamental objective to detect in a technical way, how the variations in consumption in the use of cellular telephony are out of the ordinary, considering this point as our most important objective. The filtering of this service is observed due to illegal acts of the unscrupulous users, to the detriment of the line owners; it is for this reason that through this research we try to detect filtration techniques to recommend alternative solutions.

In the development of this work the researcher reveals how information structures are constructed that present recent and historical data of each of the users, where we make use of information that contains a call, it is observed in the determination of the calls, to say if the LOC, NAT and INT calls, being complex also complex the construction of a function with so many variables of parametric inputs, which are not always known.

Since the presence of different fraud detection methods has been determined, all of them have efficient results, always yielding in incomplete quantities to those that the authors know as spikes work with consumption peaks or fixed rules, for which the applied methodology has that be according to the requirements; for the same thing it is necessary to mention that it is a basic research and as argued by Hernández, R. (2010, p. 192), Basic research is characterized by the fact that they are directed especially to distinguish and apply as the general laws of phenomena studied, a aspect that allow us to determine the most significant and discrete theories.

The solution we present has been elaborated using the technology of unsupervised neural networks, in particular the SOM networks. Among them is presented:

- The graphics where the patterns of the lists created with their respective values are observed.

- Tests of Local, National and International calls

In the same way, it is concluded in the following:

- a) The structure that the LOC, NAT, and INT call patterns should have.

- b) It has been established that the structure records for the LOC service must be of 144 patterns: NAT must be 64 patterns and INT, 64 patterns.

INDICE

Carátula	i
Dedicatoria	ii
Agradecimientos	iii
Resumen	iv
Summary	vi
Índice	viii
Introducción	xii
 II PROBLEMÁTICA DE LA INVESTIGACION	
2.1. Descripción de la realidad problemática	1
2.1.1 Actividades relacionadas con la instalación y mantenimiento de los “perfiles del usuario”.	
2.2 Formulación del problema	2
2.3. Justificación del problema	3
2.4 Hipótesis y variables	
III.- OBJETIVOS DE LA INVESTIGACIÓN	
3.1 Objetivos generales	7
3.2 Objetivos específicos	
IV. MARCO TEÓRICO	
4.1 Aspectos técnicos	8
Propuesta de solución	10
4.2 Elaboración de perfiles del usuario	11
4.3 Los cambios de comportamiento: Posible solución	15
Solución a las situaciones de acción	16

4.4	Dificultades en las propuestas de solución	16
4.5	Redes Neuronales	17
4.6	Representación del problema por resolver	18
4.7	Complemento a la descripción y definición del fraude	19
4.8	Bases teóricas	
	Fraude contractual	
	La violación de seguridad como medio de fraude	20
	Fraude técnico	21
	Fraude durante el procedimiento	22
	Soluciones propuestas	
	Perfiles de usuario: control de construcción y mantenimiento	23
4.9	Marco conceptual	28
	El aprendizaje	29
	Conocimiento sobre detección de cambios de comportamiento	
	Alternativas de solución a la magnitud del problema	30
V METODOS Y PROCEDIMIENTOS		
	Métodos empleados	31
5.1	Pruebas de creación de patrones de consumo	32
5.2	Construcción de perfiles y detección de comportamientos	
5.3	Criterios aplicados en la creación de patrones informativos	33
5.4	Tamaño de la estructuración de perfiles	34
5.5	Valores asignados a las pruebas	35
	Información empleada en la experimentación	
	Reconocimiento de llamada	36
	La hora en que se realiza la llamada	

El tiempo que dura cada llamada	
5.6 Las redes neuronales: Pruebas	37
Perfiles de usuario y detección de cambios de comportamiento	
HIPÓTESIS Y VARIABLES	
5.7 Planteamiento de Hipótesis	38
Variables	39
Variables independientes	
Variables dependientes	41
Funcionamiento de cada variable	
Independientes	43
Dependientes	
Operacionalización de Variable: detección de fraude	
en telefonía celular	44
Operacionalización de Variable: utilización de redes neuronales	45
Con respecto a la creación de patrones	46
Los perfiles como medios de cambio de comportamiento	49
El UPH actualizado con cada llamada - Hellinger	50
Sensibilidad del UPH analizado con el Umbral Hellinger	52
ESTRATEGIAS METODOLÓGICAS	
5.8 Tipo de investigación	54
Nivel de investigación	
Diseño	
Población y muestra	55
Técnicas de recolección de datos	56

5.9 Validez y confiabilidad	57
Validez	
Confiabilidad	58
VI DISCUSIÓN DE LOS RESULTADOS	61
VII CONCLUSIONES Y RECOMENDACIONES	
7.1 Conclusiones	62
7.2 Recomendaciones	64
Referencia	66
Anexos	67
 <u>ÍNDICE DE GRÁFICAS</u>	
Gráfica N° 01 Representación del problema por Resolver	18
Gráfica N° 02 Patrones de llamadas Locales (LOC)	47
Gráfica N° 03 Patrones de llamadas Nacionales (NAT)	48
Gráfica N° 04 Patrones de llamadas internacionales (INT)	49
Gráfica N° 05 Distribución de frecuencias CUP	50
Gráfica N° 06 Análisis del UPH – Frecuencias UPH	51
Gráfica N° 07 Análisis del UPH – Lanzamiento de alarma	52
Gráfica N° 08 Orden de frecuencias UPH	64

I.- INTRODUCCION

Al inicio de toda llamada de celular, las celdas o interruptores indican la utilización de éstas, produciendo la información necesaria para este caso. A este tipo de registros informativos lo denominamos CDR's (Call Detail y Records). Los CDR's en el idioma español significa Registros de detalles de llamadas, estos registran importante información sobre las llamadas, para que se le cargue la a quien corresponda. Estos mismos registros se pueden emplear en la detección de actos fraudulentos, a los cuales se les considera indicadores de fraude con tendencias avanzadas. Es decir, tienen la capacidad de analizar y procesar cantidades fuertes de CDR's recientes, haciendo comparaciones con una función de los diferentes espacios de acción, tales como IMSI (International Mobile Subscriber Identity, que traducido al español significa "Identidad Internacional del abonado de telefonía móvil, el mismo que posee la capacidad de identificar usuario que realiza llamadas mediante esta red en una red), indicando la fecha en que se realiza la llamada, el tiempo, la clase de llamada y se acciona solamente aplicando criterios de reserva. Si esta función expresa un valor que indica estar fuera de los límites normales, la alarma se activa, que el analista debe entender que se está produciendo un fraude y para constatar si realmente hubo o no actitud fraudulenta. Para el proceso de los CDR's, se necesita activar en forma previa los procesos técnicos de las telecomunicaciones, el cual está contenido en el aparato de control, en el que se puede observar una cantidad de información contenida en un formato de registro,

en él se manifiestan los CDR's (que en la práctica es conocido como un interruptor); codifica en un formato nuevo de registro, propio del sistema de fraude en este caso. Los sistemas de control conocidos en el sistema para la detección de fraude intentan consultar la secuencia a los CDR's comparando alguna función de los campos con criterios fijos conocidos como un desencadenador o Disparadores. Un desencadenador, cuando es activado, genera la alarma que el analista lo considera como fraude. Este tipo de sistemas generan resultados lo que se les conoce como medio de análisis total de CDR's; precisamente a esta actividades se les reconocen como medio de detección de los extremos de la actividad fraudulenta. Para poner en activar otras formas de análisis que marquen diferencia con este tipo análisis, por lo general se tiende a monitorear patrones de comportamiento de los teléfonos celulares, comparando las tendencias con el historial o uso historial de las llamadas más próximas, haciendo comparaciones entre sus actividades más recientes, si se observa un cambio en el patrón del comportamiento, es un indicio sospechoso, que puede ser una actividad fraudulenta (APC 2003).

Los métodos empleados para realizar fraude o robo en este tipo de servicios están constantemente cambiando, evolucionando en forma alarmante; debido a que la tecnología es cada vez más elevada y al mismo tiempo el método de control de dolo en esta materia se estanca más, generándolo mayores posibilidades al fraude. Se observa que cuando por vez primera se lanzaron al mercado, su consideración fundamental siempre se basaba en su seguridad, pero existía cierta inseguridad, manifiesta particularmente en la falta de encriptación de los datos en los canales de comunicación que permitía la clonación de teléfonos celulares, es decir, dos aparatos celulares diferentes usando la misma cuenta. Como es de conocimiento público que la tecnología evolucionó, desde el caso de ver los problemas por deducción o analogía, a una era digital, para lo cual la tecnología del robo o fraude, tenía que cambiar, es decir, la naturaleza del robo, también tenía que tecnificarse, disminuyendo las posibilidades clonación, situación que obliga a los individuos crear técnicas más avanzadas, conocido como fraude de mayor sofisticación; ante lo cual se estima que, ni las redes digitales se libraban completamente de este tipo de clonación. Las pérdidas económicas de las empresas dedicadas a este mundo

de las telecomunicaciones sumaban varios millones de dólares. Todo esto hace que los, métodos de control de este tipo servicios, es decir, la detección y prevención de la modalidad de fraude sea una de las actividades más importantes en el mercado de telecomunicaciones. En los siguientes avances de la investigación se presentan de modo algunos ciertos conceptos que serán desarrollados en profundidad a lo largo de todo el trabajo.

II.- PROBLEMÁTICA DE LA INVESTIGACION

2.1 Descripción de la realidad problemática

Cuando se trata de instalar un sistema de detección de fraude en base a los sistemas diferenciales, es preciso tener en cuenta una serie de casos que comúnmente se consideran en este tipo de investigación, los cuales lo analizamos en forma cuidadosa los aspectos o antecedentes como las siguientes:

2.1.1 Actividades relacionadas con la instalación y mantenimiento de los “perfiles del usuario”.

Por lo general los indicadores de fraude no son propios para que un investigador se detenga en esta parte, sino los aspectos más profundos haciendo uso de un único CDR. La mayoría de los fraudes, se pueden detectar con el uso de una secuencia de los mismos. Es decir, para la detección de un fraude diferencial; es necesario disponer de la información acerca de los antecedentes, aunado a las muestras de sus actividades últimas. Dentro de los intentos iniciales buscando la solución del caso, se puede extraer y codificar la información de los CDR's y mantenerlos en formato en formatos de registros especiales, para lo cual es necesario disponer de dos tipos de registro: Uno de ellos que contenga la información más reciente, llamado CUP (Current User Profile) o Perfil de Usuario de Current; y otro con la información histórica llamado en nuestro caso UPH (User Profile History), sustentado por Burge & Taylor. Cuando la información de un nuevo CDR se procesa, el acceso de mayor antigüedad del registro UPH tiende a ser descartado; pero la más antigua del CUP se puede ingresar al UPH. De este modo el nuevo registro codificado es ingresado al CUP. Por lo cual esta información debe ser conservada en forma compacta, siempre brindando las posibilidades de ser fáciles de ser analizados luego por el sistema de detección de fraude o para cualquier otro tipo de requerimientos. Considerando la gran cantidad de información que contiene un CDR es necesario encontrar otros métodos de “clasificar” las llamadas en diferentes grupos o prototipos, donde cada una debe pertenecer a un determinado grupo. Aquí se plantan interrogantes útiles para contestar: (a) ¿Cuál debe ser la estructura de los

registros CUP y UPH?, (b) ¿Cuál debe ser la cantidad de prototipos que deben tener cada uno de los registros CUP y UPH para contener y conservar la información necesaria?, (c) ¿Cuál debe ser la forma clasificar las llamadas en cada uno los diferentes prototipos definidos? y (d) ¿Cuáles deber ser los métodos de codificación de las llamadas para que sean consideradas como prototipos?

2.2 Formulación del problema

Construir un sistema de fraudulento basado en un análisis diferencial es de suma importancia plantearse situaciones problemáticas que comúnmente, en estos casos, los mismos que deben ser tratados en forma cuidadosa.

según Moreau y Preneel, en su Information escrita “Security and Cryptology - Icisc 2004”). ¿de qué manera se puede representar el consumo actual e histórico de cada uno de lao usuarios a través de los perfiles del usuario?

- se plantea métodos que nos permitan detectar cuándo y qué medida un usuario cambia su comportamiento, y tiende a discutirse estos cambios, cuantificando la utilidad obtenida por el fraude cometido en el sistema.

- Finalmente se resume y se presenta en forma gráfica el problema a resolver en su respectivo grupo o conjunto.

2.2.1 Planteamiento del problema

Problema General:

¿De qué manera se relacionan la Detección de fraude en el uso de telefonía celular con la Utilización de las Redes Neuronales en Lima ciudad de, periodo 2015?

Problemas Específicos:

- ¿Cuál es el grado de relación se relacionan entre la Detección del fraude en el uso de redes neuronales con los costos de los servicios de telefonía celular en Lima distrito, año 2015?

- ¿Qué grado de relación existe entre el uso de las redes neuronales con la estructura de patrones LOC, NAT y INT en la ciudad de Lima, periodo 2015?
- ¿Cuál es la relación que se observa entre el uso de las redes neuronales con la elaboración y mantenimiento de los perfiles del usuario en Lima distrito, periodo 2015?

2.3. Justificación del problema

En toda actividad de esta naturaleza, siempre surgen ciertas dificultades, estas dificultades en este tipo de sistemas es de bastante importancia para que los controles de detección de fraudes sean más eficientes. Se debe tener en cuenta que los procesamientos de la información se realizarán sobre millones de llamadas por día, las cuales servirán para construir los perfiles de cada uno de los usuarios de la empresa; considerando estas servirán para determinar el tamaño de la estructura, a su vez, dependiendo del tamaño de ellas se pueden generar cambios notables, tales como reducir de cientos de miles de usuarios a un par de millones. Esto indican que, que las cantidades de información que se tiene que conservar y las formas de acceder a ellas son puntos tan importantes como el mismo sistema de detección de fraude. En estos casos se tiene siempre en consideración la velocidad de procesamiento en función a la cantidad de información que debe analizarse. Una de las modalidades aplicadas en la detección de fraude que se tome tiempos en cierto modo prolongados en días en analizar una cantidad mínima de llamadas resulta ser un mecanismo bastante obsoleto, alejándose del objetivo central de la empresa.

Entonces el problema que afrontamos, basado no solamente con la intención de detectar variaciones en el consumo de llamadas, sobre todo aquellas que se observan fuera de lo normal, sino que también importa bastante los aspectos relacionados fundamentalmente con la construcción de las estructuras de información que informen el comportamiento reciente e histórico de todos y cada uno de los usuarios, teniendo en consideración la cantidad de información que contiene una llamada y la complejidad de la CACIC 2003 – Red UNCI 1338, es decir, la construcción de una función con

muchas variables de entrada, lo cual es obviamente bastante compleja y poco conocida por las tecnologías recientes.

- La detección de cambios en el comportamiento del usuario
Si ya se ha alcanzado alamar la imagen que se observa en la codificación, comparado con la información histórica de tal comportamiento de cada usuario, es preciso, adquirir unan forma de análisis de información, que detecte las anomalías en el comportamiento, donde puede disparar una alarma, que es muy probable que se trate de un fraude en el uso de telefonía celular. Generando la siguiente interrogación: ¿Cuál es la manera de detectar el cambio en el patrón o lista en el consumo?

- HIPÓTESIS Y VARIABLES

2.4 hipótesis

2.4.1 PLANTEAMIENTO DE HIPÓTESIS

Ha. Hipótesis General

Se ha determinado que existe una relación significativa entre la detección de fraude en el uso de telefonía celular con el uso de las Redes Neuronales en Lima distrito, periodo 2015.

Hipótesis Específicas

H₁. Se ha podido establecer un alto grado de relación entre el fraude en el uso de telefonía celular con los costos de los servicios de comunicación en Lima distrito, año 2015.

H₂. Existe una relación bastante considerable entre la detección de fraude en el uso de la telefonía celular y las estructuras de patrones de llamada LOC, NAY y INT en Lima distrito, año 2015.

H₃. Los estudios demuestran que existe una relación determinante entre el fraude en el uso y control con las redes neuronales y la construcción y mantenimiento de los perfiles del usuario en Lima distrito, periodo 2015.

2.4.2. VARIABLES

a) Independientes

- Las redes neuronales en los procesos donde se busca determinar en forma clasificada los diferentes tipos de llamadas, siempre requieren de dimensiones lo suficientemente grande como para poder hacer clasificaciones amplias. En cuanto a las llamadas locales indica una dimensión $NL \times ML$, que describe el Mapa de Auto Organización (SOM), que permite clasificar las llamadas. Locales.

- Disponer de la Red Neuronal con una dimensión, útil para determinar la clasificación de las llamadas Nacionales ($NN \times MN$), La cual tiene una dimensión de la matriz que hace descripciones del Mapa de Auto Organización (SOM) para clasificar las Llamadas Nacionales .

- El tamaño de la red neuronal, Útil para clasificar todas las llamadas Internacionales ($NI \times MI$), la misma que define el desarrollo de la, matriz que permite describir el Mapa de Auto Organización (SOM), en el caso del análisis de las llamadas internacionales.

- Con este mismo trato las tasas de aprendizaje estarían presentando un comportamiento estático (α): Es la misma tasa estática de aprendizaje que nos permite adaptar cada una de las neuronas (patrones) de nuestra red. El valor de α se tiene que determinar técnica experimentalmente en valores más adecuados, que oscilan entre 0,6 y 0,9. Si estos valores fueran más inferiores, demostrando la gran sensibilidad de cada nueva llamada, donde se observará los cambios en los patrones presentado en forma rápida actitud por el cual las llamadas anteriores tienden a perderse, en cierta forma distorsionando así los estudios realizados.

- Es importante definir la distancia máxima que se debe considerar entre neurona vecina con una neurona afectada (DVMAX). Los análisis en estos casos nos permiten los patrones actualizados en el entrenamiento de las redes SOM. Lo importante es hacer las definiciones y mediciones de las funciones de las neuronas vecinas, que en nuestros análisis se han convertido en dependientes de la variable DVMAX.

Es necesario recordar las reglas de aprendizaje en un SOM (Hollmen, 2008): Los analistas esperan aumentos de ingresos.

$$M_i(t_1) = m_i(t) + h_{ci}(t)(x(t) - m_i(t))$$

Donde:

$m_i(t_1)$ = Representa al peso de la neurona i en el tiempo t_1

$m_i(t)$ = Representa al peso de la neurona i en el tiempo t .

$x(t)$ = Representa al patrón de entrenamiento t (o llamada codificada).

$H_{ci}(t)$ = Representa a la función de vecindario en el tiempo t , definida como:

$$h_{ti}(g) = \alpha(g) \frac{D_{VMAX} - d(t)}{D_{VMAX}}$$

Donde:

$\alpha(t) = \alpha(1-g/gMAX)$ (con α : tasa de aprendizaje estática): tasa de aprendizaje en el tiempo g .

$d(g)$ = Distancia entre la neurona hasta la neurona vecina, es decir, a la neurona BMU en el tiempo g (A medida que el tiempo g avanza, $h_{ci}(g)$ tiende a cero; lo que indica, que el vecindario se hace más pequeño y las neuronas vecinas que se actualizan disminuyen).

b) Variables dependientes

a) La lista de patrones de uso empleados para la clasificación de las llamadas locales (PL): indica que estos patrones se utilizarán en la red neuronal para las llamadas locales, para de esta manera, el mismo que se obtiene multiplicando el NL x ML.

b) El rol de patrones necesarios para la clasificación de las llamadas DDN (PN), en lo referente al listado que se empleará en el control de la red neuronal de diferentes tipos de llamadas, para analizar todo el espacio de llamadas; valor que se calcula multiplicando NN x MN, es decir en lo referente a las llamadas nacionales.

c) En lo que respecta a los listados utilizados en la clasificación de las llamadas con destino Internacional (PI), relacionado con la cantidad de patrones de los usuarios que emplearemos en la red neuronal para las todas las llamadas Internacionales para clasificar todo el espacio de

llamadas en un tiempo de análisis previamente establecido, obteniéndose al operar NI x MI,

d) En que respecta a las dimensiones de los perfiles CUP y UPH (K): En los cuales se indican la cantidad total de patrones, que contienen los perfiles de los PL patrones LOC, PN patrones NAT y PI patrones INT.

III.- OBJETIVOS DE LA INVESTIGACIÓN

3.1 Objetivos generales

Determinar en forma exacta la relación existente entre la detección de fraude en el uso telefonía celular con la utilización de las Redes Neuronales en Lima distrito, año 2015.

3.2 Objetivos específicos:

- Establecer la relación que mantienen el fraude en el uso de las redes neuronales con los costos de los servicios de telefonía en Lima distrito, periodo 2015.
- Establecer la relación en influencia entre la detección de fraude en el uso de las redes neuronales y las estructuras de patrones de llamadas LOC, NAT e INT en Lima distrito, año 2015.
- Establecer la relación e influencia entre el fraude en el uso de las redes neuronales y la elaboración y mantenimiento de los perfiles del usuario en Lima distrito, año 2015.

IV. MARCO TEÓRICO

4.1 ASPECTOS TÉCNICOS

La estructuración, construcción y el mantenimiento de los perfiles en Lima distrito:

En un contexto social la mayor parte de los indicadores de fraude tienden a analizarse de diferentes maneras, es decir, los CDR tienen muchas formas para ser analizados; una gran parte de los fraudes pueden ser detectados mediante una secuencia de los mismos, mencionados en su libro "Detección de fraude en telecomunicaciones utilizando la divergencia" (Taniguchi, Haft, Hollmen & Tresp, 2008). Cuando se hacen los análisis mediante un sistema de detección de fraude diferencial, se requiere información acerca del consumo histórico a las muestras de su actividad más reciente; es decir, un conjunto de CDR's con la suficiente cantidad de llamadas que puedan caractericen a un usuario. Una acción inicial podría ser extraer y codificar la información que contienen los CDR's y conservarla en un determinado formato de registro (ASPeCT, 2000). Todo esto indica que en una situación inicial se requieren dos tipos de registro: uno que almacene la información más reciente, al que en este trabajo lo denominamos CUP (Current User Profile, que en español significa: Perfil del Usuario Actual) y un segundo la que conserve la información histórica, al que denominaremos UPH (User Profile History, que en español significa; Historial de Perfil de Usuario), estudiado por Burge & Taylor en 2004. Ellos decían que. Cada vez que un nuevo CDR de un determinado usuario ingresa para ser procesado, la entrada más antigua del registro UPH debería ser descartada y la más antigua del CUP debería ingresar al UPH. Entonces este nuevo registro codificado, debería ingresar al CUP en su libro "Prueba de la capacidad de

detección de fraude de diferentes perfiles de usuario mediante clasificadores” (Gosset & Hyland, 2005).

La información almacenada y conservada en forma compacta y con características fáciles posteriormente por el sistema de detección de fraude. Teniendo en cuenta la gran cantidad de información que contiene un CDR, es necesario conseguir un método de fácil clasificación, estas deben estar en las llamadas en grupos n o prototipos, donde cada una de éstas debe pertenecer a un único grupo. Es decir, que bajo estas condiciones se plantean diferentes cuestiones que son muy importantes, las que se deben resolver:

- a) ¿Cuál es el modelo de estructura que deben tener los registros CUP y UPH?
- b) ¿Cuál es el número de grupos o prototipos deben poseer los registros CUP y UPH para contener la información necesaria?
- c) ¿De que manera se pueden clasificar las llamadas en los diferentes prototipos definidos?
- d) ¿De que manera se deben codificar las llamadas para que estas puedan convertirse en prototipos?
- e) La detección de cambios en el usuario, manifestado en el comportamiento de los usuarios.
- f) Después de la construcción de una imagen codificada del consumo reciente y la historia de cada usuario del usuario, la información es analizada al detalle y observar en forma minuciosa para detectar las anomalías, objetivo del estudio. Luego de todo esto es preciso plantear una segunda cuestión:
 - g) ¿De qué manera se puede detectar el cambio en el patrón de consumo de un individuo?
 - h) La dificultad de la acción en un determinado caso.
 - i) Los actos de detección y control en este tipo de sistemas son bastante polémicas. Al respecto Seymour, (2000); decía: que se debe tener en cuenta que se procesarán millones de llamadas por día, las mismas que

serán utilizadas para construir los perfiles de cada uno de los usuarios de la empresa, las cuales, dependiendo de la magnitud de la misma, puede variar de cientos de millones a un par de millones, Symour sostenía lo que ya lo habían hecho en (1999) Frank, Hunt & Davey, (2002). Es decir, que la cantidad de información a acumular y la forma de acceder a esta información, son puntos de gran importantes como el sistema de detección de fraude (Fawcett & Provost, 2009) en sus apuntes

- j) " Diseño automático de sistemas de detección de fraude ". También es importante considerar la velocidad de procesamiento que se debe ejecutar, pero resulta ser este caso muy relativo, debido a la cantidad de información que debe procesarse. En un sistema de detección de fraude que demore días en ser analizados, más aun cuando se trata de pocas llamadas es considerado totalmente anti técnico y no cumple con los objetivos en tiempos requeridos por las compañías.

Obviamente el trabajo se realiza en Lima, pero por la amplitud del sistema, es un problema sin fronteras y también las llamadas se realizan a todas partes del mundo. En base a lo que enfocamos el problema, se sostiene que esta problemática lo estamos direccionando, específicamente basado en la detección de cambios de consumo fuera de lo normal, en la construcción de estructuras completas de información que representen el comportamiento reciente e histórico de cada uno de los usuarios, considerando la gran cantidad de información que contiene cada llamada y la complejidad de la construcción de la aproximación a una función con tantas variables de entradas desconocidas.

Propuestas de solución

Ante la descripción del problema planteado, es preciso de proponer algunas formas de solución, tal como lo expresamos a continuación:

- a. En esta parte se define el formato de registro necesario. Es decir, se detalla la salida de la mediación.

- b. En seguida se debe detallar la forma cómo se construyen y se conservan los perfiles con la utilización de las redes neuronales SOM.
- c. En este punto se define una distancia vectorial que sirve para determinar qué tan diferentes son los perfiles CUP y UPH de un usuario.
- d. En esta sección se debe describir en forma detallada el procesamiento de la información como una manera de minimizar los problemas objeto del estudio.
- e. En esta parte del planteamiento se mencionan algunas limitaciones que se pueden anteponer a las soluciones.
- f. Análisis y procesamiento de la información – Mediación

Inicialmente al momento del procesamiento de los CDR's se debe crear un formato de registro con información que identifique al usuario (el IMSI es la información adecuada para dicho fin); la fecha de la llamada en un formato fijo que incluya cuatro dígitos para el año, dos dígitos para el mes y dos dígitos para el día, al que se le definirá con la sigla AAAAMMDD; la hora de la llamada en otro formato fijo que contenga 2 dígitos que indique la hora (00 a 239), 2 dígitos que indique los minutos y otros dos para los segundos al que lo definiremos con la sigla **HH24MISS**; donde es con facilidad se controla la duración de la llamada en segundos en los formatos que empleen cinco dígitos con ceros a la izquierda si corresponde a un tipo de llamada clasificada en LOC (llamada local), NAT (llamada DDN o nacional) e INT (llamada DDI o internacional); este formato de registro será el resultado del proceso de mediación. Con esta información procesada y acotada a los datos necesarios, se pueden en gran forma responder a muchas cuestiones en lo concerniente a esta materia, donde las de mayor importancia son cuestionarios sobre la información de entrada y la información de salida de la mediación. I

4.2 Elaboración de perfiles del usuario

Lo primero que se tiene que resolver es a través de interrogante es cómo se construirá los perfiles CUP y UPH; es decir, la determinación de los patrones que compondrán cada uno de estos perfiles. Estos patrones deberán contener información del consumo del usuario, especificando el tipo de consumo LOC, NAT e INT respectivamente. Una de las maneras de cómo construir estos

patrones es haciendo uso de las redes neuronales (ASPECTO, 1997), para que se considere con mayor discreción es usando el espacio de todas las llamadas de los usuarios, es decir los tiempos entre llamada y llamada generando un espacio de "n" patrones que representen el consumo de todos los usuarios y luego creando una distribución de frecuencias por cada usuario, en donde se represente aspectos de probabilidad de hacer llamadas de ese patrón disponible para un usuario. Resumiendo lo sostenido, cuando se elabora el perfil del usuario, lo que se hace es representar la distribución de frecuencias, donde un usuario cualquiera que sea, realiza un determinado tipo de llamada, mostrándose en esta estructura de información el patrón de consumo del mismo.

Se debe analizar la importancia de las redes neuronales, las mismas que, ante tantas ventajas observadas en el mundo de las telecomunicaciones, tienen importancia en los casos de clasificación de la información en determinados patrones (Hilera González & Martínez Hernando, 2000); especialmente las redes SOM (Self Organizing Map, que en español significa: Mapa de Auto Organización, que permiten al investigador adquirir la información necesaria para construir estos patrones de manera no supervisada por medio de criterios de semejanza, enfocado por Kohonen (2000). En el caso que corresponde a la siguiente investigación, es posible procesar todas las llamadas realizadas por todos los usuarios para que las redes generen los patrones, según sea la cantidad que hay de cada tipo, pero que estos patrones representen a todas ellas. En este tipo de procesos, para evitar ruidos molestos que puedan entorpecer la captación de la información, se utilizan 3 redes neuronales que generen patrones para representar a las llamadas LOC, NAT e INT respectivamente; el perfil del usuario se elabora haciendo uso de todos los patrones creados por las 3 redes. Un tipo de información que se emplean para representar un patrón son la hora en que se realiza la llamada y la duración de la misma; considerando que si lo representamos en un eje cartesiano la hora de todas las llamadas y la duración correspondiente, se obtiene un rectángulo prácticamente lleno de puntos. La idea en este tipo de análisis es obtener un gráfico en el cual sólo aparezcan los puntos más representativos de todo el espacio en cuestión; precisamente esta es la tarea de las redes neuronales. Además este diseño de 3 redes neuronales nos permite, no solamente detectar

cambios de comportamiento, sino que también representa de manera general el comportamiento de todos los usuarios de la empresa; es decir, que observando los patrones generados por cada una de las redes neuronales en un gráfico, será posible con facilidad obtener las conclusiones donde se muestren la manera de cómo se comportan en general los usuarios de la empresa y tomar las decisiones del tipo comercial, agregando una funcionalidad, que no solamente se detecte el fraude y se soluciones los problemas, sino también las ventajas que se pueden obtener.

En todos los puntos que se está desarrollando en este trabajo de investigación, se amplían y remarcan los conceptos más interesantes. Una vez obtenidos los patrones necesarios que se emplearán para representar los perfiles del usuario, es imperioso comenzar con el llenado de la información en los formatos; el procedimiento consiste en captar la llamada que se va analizar, codificarla y que la red neuronal determine a qué patrón se parece más dicha llamada. Una vez obtenida la información requerida, se debe adaptar al perfil de usuario CUP, de manera que la distribución de frecuencias, muestre que el usuario tiene ahora una mayor probabilidad de realizar este tipo de llamadas. Sabiendo que el perfil de usuario tiene K patrones que se componen de **L patrones LOC, N patrones NAT e I patrones INT**, se puede construir un perfil bastante representativo de la llamada procesada y luego adaptar el perfil CUP con dicha llamada. Si la llamada es LOC, los N patrones NAT y los I patrones INT tendrán una distribución de frecuencias igual a 0, y los K patrones LOC tendrán una distribución de frecuencia dada por la ecuación de predicción de campos estocásticos en redes SOM (Grabec & Mandelj, 1998):

$$Vi = \frac{e^{-|x-Qi|}}{\sum_{j=1}^L e^{-|x-Qj|}}$$

Donde:

X: Representa a la llamada a procesar codificada.

V_i : Representa a la probabilidad que la llamada X sea del patrón i .

Q_i : Representa al patrón i generado por la red neuronal LOC.

Considere que:

$$\sum_{j=1}^L V_j = 1$$

En cuanto a lo que significa cada uno de los componentes de esta ecuación se explica de la manera siguiente:

Si se tratara de una llamada NAT, se indica que se debe reemplazar L por N , donde la distribución de frecuencias LOC e INT serían equivalentes a 0; si la llamada fuese INT, se indica que se debe reemplazar L por I y la distribución de frecuencias LOC e NAT serían equivalente a 0.

En base a lo expresado se puede definir el vector representativo de la llamada V , de dimensión K como:

$$V_i = v_i, \text{ cuando } 1 \leq i \leq L$$

$$V_i = 0, \text{ cuando } L+1 \leq i \leq K, \text{ Si la llamada es LOC.}$$

$$V_i = v_i, \text{ cuando } L+1 \leq i \leq L+N$$

$$V_i = 0, \text{ cuando } 1 \leq i \leq L \text{ y } L+N \leq i \leq K, \text{ Si la llamada es NAT.}$$

$$V_i = v_i, \text{ cuando } L+N+1 \leq i \leq K \quad V_i = 0, \text{ con } 1 \leq i \leq L+N, \text{ Si la llamada es INT.}$$

Luego de todo lo explicado, ahora que se ha definido el vector V , se puede incorporar el vector CUP a la información de la llamada procesada:

$$CUP_i = \alpha_{NAT} CUP_i - (1 - \alpha_{LOC}) V_i, \text{ con } 1 \leq i \leq k ;$$

Cuando la llamado es LOC

$$CUP_i = \alpha_{NAT} CUP_i - (1 - \alpha_{NAT}) V_i, \text{ con } 1 \leq i \leq k ;$$

Si la llamado es NAT

$$CUP_i = \alpha_{INT} CUP_i - (1 - \alpha_{INT}) V_i, \text{ cuando } 1 \leq i \leq k$$

Si la llamado es INT

Explicación de los componentes:

α_{LOC} : Representa a la tasa de adaptabilidad aplicada, cuando la llamada X es incorporada al CUP, siempre que X corresponda a una llamada LOC.

α_{NAT} : Representa a la tasa de adaptación operada, siempre que la llamada X se incorpore al CUP, si X corresponda a una llamada NAT.

α_{INT} : Representa a la tasa de adaptación calculado cuando la llamada X es incorporada al CUP, si X analizara a una llamada INT.

Una vez que se ha logrado adaptar el perfil CUP, se hace una comparación con el perfil UPH para determinar si se ha producido un cambio significativo en el comportamiento (a esto se le considera como el motor de detección de cambios de comportamiento); asimismo se entiende que una vez realizada esta tarea, queda se adaptado el UPH con la información del CUP, siempre y cuando la cantidad de llamadas necesarias para generar cambios en el patrón histórico se hayan procesado dentro de lo normal:

$$UPH_i = BUPH_i + (1 - \beta) CUP_i, \text{ con } 1 \leq i \leq k$$

Donde:

β : Representa a la tasa de adaptación aplicada, si el CUP es incorporado al UPH.

Se está realizando el análisis de los casos 1, 3 y 4, debido a que se ha definido las estructuras de los perfiles CUP y UPH y se describe la forma en que se

construirán. El caso 2, se analizará en forma más detenida en la cuarta parte de este trabajo, es decir, se resolverá.

4.3 Los cambios de comportamiento: posible solución

Para que el investigador pueda determinar si hubo o no cambios en el comportamiento del consumidor, tiene que comparar los perfiles CUP y UPH y determinar si la diferencia entre los vectores es lo suficientemente considerable como para lanzar una alarma. Debido a que el CUP y el UPH son dos vectores que representan a las distribuciones de frecuencias, podemos emplear una distancia vectorial para comparar qué tan diferentes son. Para procesar este caso se puede utilizar la distancia Hellinger (H), cuyo valor indica la diferencia entre dos distribuciones de frecuencias (Burge & Taylor, 1997). Esta distancia siempre será un valor que se encuentren entre 0 y 2, en la que cero se dispone para distribuciones iguales y dos representa a la ortogonalidad. El valor de la distancia representado por H sirve para determinar las diferencias frecuencias CUP y UPH para el lanzamiento de la alarma. Los valores de la distancia para el lanzamiento de las alarmas se miden mediante la siguiente ecuación:

$$H = \sum_{i=1}^k (\sqrt{CUP_i} - \sqrt{UPH_i})^2$$

En cuanto al caso 5 que se encuentra representada en esta ecuación busca que definir la forma de detectar las diferencias entre el comportamiento reciente y el comportamiento histórico, lo cual sirve para lanzar las alarmas, siempre y cuando corresponda al caso.

Solución a las situaciones de acción

La acción que se plantee estará sujeta en forma directa del Hardware, en la que compitan el sistema de detección de fraude y los cambios de comportamiento de los usuarios. Desde el punto de vista del software se hace

uso lo menos posible con bases de datos relacionales; y se debe hacer todo el procesamiento empleando archivos de información, con la mínima cantidad de escrituras y lecturas de disco. Es de suma importancia la comprensión e interpretación de la información mencionada, debido a que el espacio es otra restricción que debe ser prevista. Por lo cual, en la solución propuesta se está trabajando con archivos planos y se trata de conservar un archivo por cada usuario con sus respectiva información de las distribuciones CUP y UPH, así mismo la última llamada procesada y la cantidad total de llamadas procesadas por el sistema.

4.4 Dificultades en las propuestas de solución

La determinación de las soluciones que se van a proponer se realizan exactamente igual a la que se observa en el análisis diferencial del consumo del cliente. Una de las maneras que no se logre determinar se encuentran en aquellos casos donde el usuario ejecuta muchas llamadas con igual o similares características con altos consumos, generando confusión para su detección, donde su patrón de comportamiento nunca variaría. Es por esta razón que siempre es necesario hacer combinaciones en las soluciones, para mantener un sistema de detección de fraude que realice las exploraciones de los diferentes tipos de fraude. En tal sentido, es imperativo considerar el análisis absoluto como la solución posible.

Otra de las restricciones se puede notar en el caso en que los patrones de uso son estáticos, donde es fácil notar, través de los cambios de comportamiento de la empresa en forma total; ante el mismo, es de suma importancia hacer re-encontrar las redes neuronales para así determinar nuevos patrones de usuarios que completen el espacio total de llamadas y renovar la construcción de los perfiles CUP y UPH con las nuevas distribuciones adquiridas.

4.5. Redes neuronales: concepto

Son instrumentos compuestos por elementos simples interconectadas en forma masiva en paralelo y con organización jerárquica; estas interactúan con

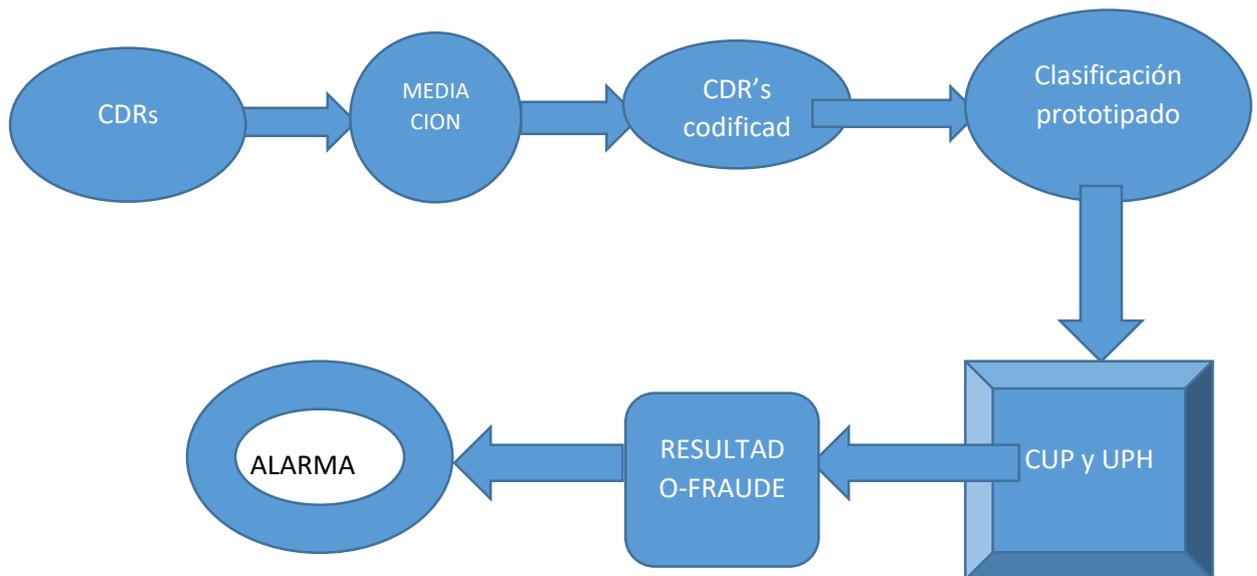
los elementos del mundo real, tal como funciona el sistema nervioso biológico (Kohonen, 1988). La complejidad observada en las operaciones de las redes es el resultado de los abundantes lazos de realimentación, junto con la poca o nula linealidad de los elementos de proceso y cambios de adaptación de sus parámetros, que tienen la capacidad, inclusive, de definir hasta fenómenos dinámicos muy complicados. (Hilera González & Martínez Hernando, 2000).

Debido a la forma como se constituye y a sus fundamentaciones, las redes neuronales artificiales presentan muchas características, bastante semejantes a la del cerebro humano. Dentro de sus funciones, por ejemplo, tienen la capacidad de aprender de la experiencia, de generalizar de casos o datos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante (Hilera González & Martínez Hernando, 2000). Entre las ventajas de mayor importancia que ofrecen son:

- Aprendizaje adaptativo: que se manifiesta en la capacidad de aprender a resolver trabajos basados en ciertas experiencias iniciales.
- La auto organización: es decir, que es capaz de crear una organización propia o representar una determinada información según los medios que pueda crear.
- Tolerancia a fallos. Se entiende que una avería parcial de una red genera una degradación de su estructura, pero algunas partes de la estructura que no sea afectada pueden retener información, inclusive si esta haya sufrido ciertos daños.
- Operación en tiempo real: Las computadoras neuronales pueden ser operados en paralelo, y se diseñan y producen máquinas con hardware especial para obtener esta capacidad (Maren, 2000).

4.6 REPRESENTACIÓN DEL PROBLEMA POR RESOLVER

Desde el punto de vista gráfico y habiendo analizado todos los problemas planteados, se presenta un resumen gráfico que contiene el sistema de detección de fraude en función a los resultados obtenidos.



Propuesta en forma gráfica

Gráfica N° 01

4.7 Complemento a la descripción y definición del fraude

Partimos de la cuestión: ¿Qué es el fraude? Desde el punto de vista terminológico se puede describir desde el punto de vista simple como “cualquier actividad dolosa, es decir, apropiarse de bienes ajenos sin el consentimiento de su propietario y sin que exista la voluntad de retribuirlo a sus propietarios. Desde el punto de vista comercial se entiende que las organizaciones hacen cálculos de la cantidad de dinero que pierden debido al robo definiéndolo como “el dinero que se pierde en clientes/cuentas por los cuales no se percibe pago alguno (Gosset & Hyland, 1999). Sin embargo, para nuestro interés, o lo que se trata de los fines de detección de dolo, robo, o lo que en estos términos la llamamos fraude. En función a este interés, la definición no es apropiada, debido que el fraude solo puede ser detectado una vez que ha ocurrido. En efecto, definir en forma clara el robo puede ser muy difícil, en función a que se basa en la diferencia entre un comportamiento fraudulento y otro que no lo es, o que la diferencia puede ser muy pequeña. Por estas

consideraciones, lo más prudente es clasificar al robo en forma diferente y hacer descripciones de cada uno de ellos.

4.8 BASES TEÓRICAS

a. Veamos cómo se puede clasificar el fraude telefónico

A continuación presentamos los diferentes tipos de fraude que deben ser considerados al momento de analizar los problemas de esta naturaleza:

- **Fraude contractual**

Todos los tipos de fraudes en estos análisis generan a priori ganancias para la empresa a través del uso de los servicios de la telefonía celular, pero observándolo desde el punto de vista de la conveniencia los usuarios, en su mayoría no tienen la voluntad de pagar los servicios, si no exista presión de las empresas que ofertan dicho servicio. Una de las maneras de cómo justificar esta afirmación es presentando en forma de ejemplo es el denominado por suscripción. El mismo puede tomar varias formas, pero puede ser dividido principalmente en dos partes: 1) Se refiere a que cuando un individuo contrata este tipo de servicios de una empresa, pero premeditadamente planea no pagar nunca el servicio consumido; 2) El segundo caso también es característico en muchos consumidores, que luego de haber acumulado varias facturaciones el usuario toma la decisión de no pagar por el uso del mismo (Cuenta acumulada). El comportamiento manifestado por el usuario genera desequilibrios en las empresas ofertantes, tal es así, que en los últimos tiempos, solamente se permiten acumular dos facturas; y sólo se repone el servicio previa cancelación de la cuenta. Estos dos casos mencionados se emplearán como modelo de control de los servicios telefónicos otorgados. En cuanto al primer caso, se observa que es un tanto dificultoso de poder ser detectado a través de información de uso, debido a que no existe información de cuándo comienza a utilizar el servicio y además es necesario disponer de información adicional, como puede ser su condición crediticia para analizar el riesgo que implique venderle el servicio a un determinado usuario.

- **La violación de seguridad como medio de Fraude**

Todas las modalidades de fraude son actos cometidos por individuos que haciendo uso de ciertas habilidades cometen actos que se ubican fuera del marco de la Ley. Estos individuos analizan y observan los puntos críticos o inseguros para cometer sus actos que se convierten en delictivos, siempre que en algún momento del futuro cancele el consumo. Es decir, utilizar los servicios de la compañía de manera ilegal. La central de la Empresa Telefónica del Perú que provee servicios de telecomunicaciones a diferentes servicios de este rubro con conexión a otras redes de telecomunicaciones (ITS, 2005), el cual nos permite visualizar que el perjuicio generado es a la red.

Cuando al usuario que comete este tipo de actos se le considera como un atacante, en este tipo de robos el "atacante" llama repetidamente al mismo número, tratando de lograr acceder y luego de lograr este acceso, tratan de realizar llamadas salientes de alto valor (nacionales o internacionales de larga duración), pagando a la Empresa proveedora las tarifas mínimas, pero en la realidad el servicio consumido tiene un alto valor. Usualmente se ha determinado que tales ataques están asociados con el uso de teléfonos clonados, es decir, los inescrupulosos de los servicios, no pagan nada, por los aparatos empleados son clonados.

Cuando se trata de ataques a la red, la intención es intentar ingresar a las redes de computadoras a través de módems que se configuran en las mismas máquinas para lograr operar trabajos más remotos de administración y soporte. Una vez que se logra acceder por uno de ellos, el atacante intenta ingresar a la red, de esta manera puede configurar ciertos equipos para su propio beneficio. En este caso los fraudes tienen la característica de componerse de ser llamadas cortas y continuas al mismo número o llamadas cortas a números secuenciales en el caso de fraude de red. Todo esto permite a que este comportamiento sea detectado.

- **Fraude técnico**

Los actos de fraudes dados de esta naturaleza se caracterizan por generar efectos que superan las capacidades tecnológicas de las Empresas de telefonía celular (Móvil). Para cometer estas acciones los individuos siempre requieren de ciertos conocimientos y habilidades. A su vez, es necesario también reconocer, que en cierto modo, las mismas empresas son culpables por no disponer de medios de seguridad de la información, la misma fácilmente se les distribuye a los usuarios sin los conocimientos necesarios que puedan emplearlo. Los ejemplos de este tipo de fraude se encuentran en la clonación de teléfonos y el fraude interno técnico. En los casos de una clonación, los parámetros de autenticación de un móvil son copiados a otro equipo, de modo que la red genera acceso de llamada para el teléfono original quien está intentando validarse.

Los trabajadores de las empresas telefónicas pueden generar cierta alteración en la información a través de los equipos. En este tipo de fraude, trabajadores de la empresa de telefonía pueden alterar cierta cantidad de información en los equipos de comunicaciones para permitir a determinados clientes reduzcan el costo de acceso a los diferentes servicios. El comportamiento de uso de estos clientes depende de cuánto tiempo desean permanecer sin ser detectados. En la situación en la que el infractor cree que puede pasar desapercibido por un buen tiempo, deberá permanecer o mantener el comportamiento normal de uso para no ser detectado. Si cambiara su estilo de uso, la solución propuesta en este trabajo lo encontraría rápidamente. Generalmente este tipo de fraude no tienen mucha duración, esto se debe a que se intenta hacer uso del servicio lo máximo posible hasta ser detectado y cortado el servicio.

- **Fraude durante el procedimiento**

Las modalidades de fraude que se consideran y se relacionan en este trabajo se realizan con la intención crear estrategias técnicas para detener este acto doloso. Siempre este tipo de desavenencias se presentan cuando

se observa debilidad en los procedimientos técnicos, con la finalidad de ingresar a los sistemas.

Como una forma típica de conocer este tipo de fraude es el de Roaming (que consiste en utilizar el teléfono en otra red, por ejemplo, otro país, y cobrarse el consumo en el país de origen). Dentro de esta modalidad, el procedimiento de facturación generalmente se realiza en tiempos posteriores, a veces prolongados, al tiempo en que realmente ocurrió las llamadas, inclusive cuando la persona que cometió el fraude no se encuentra dentro del plano geográfico indicado.

a. Soluciones propuestas

La solución que se está proponiendo y desarrollando, se ha considerado cada una de los casos planteados en forma anticipada, con la intención de realizarlo eficazmente posible. Es así que nos permitimos mencionar a continuación cada una de las posibles respuestas, en basado en las interrogantes que se hicieron con relación al problema. Todas estas implicancias indican que para poder comenzar a procesar los CDR's es necesario disponer de un formato de registro (salida de la mediación) que contenga la información siguiente: IMSI (identifican del usuario), decir, la fecha de la llamada en formato AAAAMMDD, la hora de la llamada en formato HH24MISS, la duración de la llamada en formato 00000 y el tipo de llamada clasificada en LOC (llamada local), NAT (llamada DDN o nacional) e INT (llamada DDI o internacional). Con la disponibilidad de la información requerida, se puede iniciar a la solución de las interrogantes planteadas, haciendo uso de los datos de entrada y salida.

- **Los perfiles del usuario: Control de construcción y mantenimiento**

El primer caso urgente de solucionar es la determinación de cómo elaborar los perfiles CUP y UPH. Es decir, cómo podemos determinar los patrones que en el futuro constituirán el perfil de los usuarios. Estos patrones deben contener la toda la información del consumo que realiza el usuario por separado, es decir, información del consumo LOC (llamadas locales), NAT (llamadas nacionales o DDN) e INT (llamadas internacionales o DDI),

respectivamente. Para lograr este tipo de herramientas de control, es por medio del uso de redes neuronales, indicando que se refiere a procesos llevados a cabo en forma discreta y haciendo uso de los espacios de todas las llamadas de los usuarios, generando así un espacio común de n patrones que a su vez representen el consumo de todos los usuarios y para posteriormente generar una distribución de frecuencias por cada uno de los usuarios, en la que se represente, qué probabilidad existe de hacer llamadas del patrón tiene cada usuario. En conclusión, cuando se tiene la oportunidad de lograr construir el perfil del usuario, se estará representando la distribución de frecuencias, en la que se puede notar que un determinado usuario realiza un tipo de llamadas determinada, mostrando en esta estructura de datos el patrón de consumo del usuario. Entre otros casos se puede considerar las ventajas que genera; las redes neuronales tienen las características de clasificar la información en determinados patrones. Especialmente, se trata de las redes SOM (Self Organizing Map), este tipo de instrumentos pueden tomar la información y construir los patrones de manera no supervisada por otros investigadores de semejanza, sin establecer nada a priori de los datos (Hollmen, 1996). En nuestro caso, con los elementos que estamos mencionando se pueden procesar todas las llamadas realizadas por todos los usuarios para que las redes, según la cantidad que hay de cada tipo, pueda generar los patrones (formando grupos de semejanza) que representen íntegramente a todas a todas ellas. En tales situaciones donde siempre no falta ruidos en la captación de información, se suelen utilizar 3 redes neuronales que tengan la capacidad de crear patrones capaces de identificar las llamadas LOC, NAT e INT respectivamente. Como es de conocimiento el perfil del usuario se construye haciendo uso de todos los patrones generados por las 3 redes. Las informaciones empleadas para representar un patrón son: la hora y duración de la llamada. Se sabe que si representamos en un eje cartesiano la hora de toda la cantidad de llamadas y la duración correspondiente, se obtendrá un rectángulo lleno de puntos; pero la idea es obtener un gráfico donde solamente aparezcan los puntos más representativos de todo el espacio en cuestión; precisamente son estas la tarea de las redes neuronales. Una vez que se haya obtenido obtenidos los patrones que se utilizarán para identificar los perfiles del usuario, es

conveniente comenzar llenar de información dichos patrones. El proceso de este sistema consiste en adquirir la o las llamadas que se desea analizar, se codifica para que la que la red neuronal determine, a qué tipo de patrón se parece la misma. De la misma forma, una vez obtenida la información requerida, se debe adaptar el perfil de usuario CUP de manera que la distribución de frecuencias, muestre tiene ahora una probabilidad mayor de ejecutar este tipo de llamadas. Enterados que el perfil del usuario posee K patrones que se componen de L patrones LOC, N patrones NAT y los I patrones INT, se pueden construir un perfil bastante representativo de la llamada procesada, para luego adaptar el perfil CUP con esa llamada. Si la llamada es LOC, los N patrones NAT y los I patrones INT tendrán una distribución de frecuencia igual a 0, y los K patrones LOC tendrán una distribución de frecuencia dada por la ecuación [Burge & Taylor, en el año 2000] y se representa con la siguiente ecuación:

$$V_i = \frac{e^{-[x - Q_i]}}{\sum_{l=1}^K e^{-(x - Q_l)}}$$

Donde:

X = Llamada procesada y codificada

V_i = Probabilidad que X sea el patrón i

Q_i = Patrón i generado por la red neuronal LOC

$$\sum_{j=1}^K v_j = 1$$

Si la llamada tuviera destino nacional interna o (NAT), indica que es importante reemplazar L por N, determinándose la distribución de frecuencias LOC e INT = 0; Si esta misma llamada fuese INT, entonces para nuestra comodidad es útil reemplazar L por I y la distribución de frecuencias LOC e NAT serán 0. En función a lo establecido en el análisis se puede definir al vector representativo de la llamada V, con una dimensión K, y expresarlo como:

$$V_i = v_i, \text{ Siempre que: } 1 \leq i \leq L$$

$$V_i = 0, \text{ Siempre que: } L+1 \leq i \leq K, \text{ Si la llamada es LOC}$$

$$V_i = v_i, \text{ Siempre que } L+1 \leq i \leq L+N$$

$$V_i = 0, \text{ Siempre que } 1 \leq i \leq L \text{ y } L+N \leq i \leq K, \text{ Siempre que llamada sea NAT.}$$

$$V_i = v_i, \text{ Siempre que } L+N+1 \leq i \leq K$$

$$V_i = 0, \text{ con } 1 \leq i \leq L+N, \text{ cuando la llamada es INT.}$$

En efecto, luego de disponer del vector V, es posible adaptar el vector CUP a la información procesada previamente.

$$CUP_i = \alpha_{LOC} CUP_i - (1 - \alpha_{LOC} V_i); \text{ cuando } 1 \leq i \leq K, \text{ Siempre que la llamada sea LOC,}$$

$$CUP_i = \alpha_{NAT} CUP_i - (1 - \alpha_{NAT} V_i); \text{ cuando } 1 \leq i \leq K, \text{ Siempre que la llamada sea NAT,}$$

$$CUP_i = \alpha_{INT} CUP_i - (1 - \alpha_{INT} V_i); \text{ cuando } 1 \leq i \leq K, \text{ Siempre que la llamada sea INT}$$

En esta α_{LOC} : Representa a la % de adaptación; generalmente se presenta cuando la llamada ingresa al CUP, si la llamada X es LOC.

α_{NAT} : Representa a la % de adaptación, Sí la llamada realizada (X) es incorporada al CUP, siempre que X corresponde a una llamada NAT.

En α_{INT} : Indica la % de adaptación, si la llamada X (Cualquiera que sea) ingresa al CUP, siempre que esta llamada X corresponde a una llamada INT.

Luego de haber SIDO adaptado el perfil CUP, es necesario hacer una comparación con el perfil UPH y se logra determinar si se ha producido un cambio lo necesariamente significativo, es decir, los cambios en el comportamiento (es un instrumento técnico útiles para determinar los cambios de comportamiento). Luego de haber concluido esta parte del análisis, se hace la adaptación del UPH con la información dada en el CUP, siempre y cuando el número de llamadas útiles para cambiar el patrón histórico han sido procesados adecuadamente, la misma que se muestra en la siguiente ecuación:

$$UPH_i = \beta UPH_i + (1 - \beta)^* CUP_i, \text{ siempre que } 1 \leq i \leq k$$

β , representa a la % de adaptación, siempre que el CUP sea incorporado en el UPH

Pero es necesario resumir la manera como se presentan este tipo de dolos en telefonía celular, tal como lo expresamos a continuación:

En los casos donde se inicia una llamada de celular, las celdas establecidas o interruptores indican y registran que dicha llamada se está realizando en el momento y generan la información referida a este acto. A este tipo de registros siempre lo denominamos CDR's (Registros de detalles de llamadas); los mismos que contienen, para posteriormente ser facturada y cobrada por la Empresa a quien corresponda. (APEC – 1996).

Los registros que estamos mencionando también pueden ser usados para detectar actividades fraudulentas, por lo cual se les conoce también como indicadores de fraude bastante sofisticados. Los procedimientos de análisis son simples y comunes en una cantidad de CDR's recientes y haciendo comparaciones con una función de los diferentes campos tales como la Identidad de Abandono Móvil Internacional, lo que en español significa Identidad de Suscriptor Móvil Internacional, que identifica en forma unívoca a los usuarios

en una red de telefonía celular, en cuanto a las fechas de llamadas , hora duración de llamada, tipo de llamada bajo un criterio de análisis (Moreau y Premeel, 1997). Según Moreau y Premeel, si esta función devuelve un valor, considerado que se encuentra fuera de los límites normales, la alarma se activa inmediatamente, que deben ser considerados por los analistas de Fraude, para verificar si realmente hubo o no actividad de mala fe.

Para ejecutar el procesamiento de los CDR's es muy importante realizar en forma previa una operación conocida en telecomunicaciones como mediación, en el mismo que se puede leer la información con reconocida en el formato de registro, donde vienen los CDR's (Estos pueden ser de longitud variable, dependiendo del tipo de llamada y proveedor del interruptor). Dado todo esto se codifica en un nuevo formato de registro más claro dentro del sistema de fraude en este caso.

Existen muchos sistemas de detección de fraude. Todos ellos intentan hacer consultas sobre las secuencias de CDR's comparando alguna función de algunos mecanismos prácticos, conocidos más, como criterios fijos, a los cuales también los llaman disparadores. Si un disparador es activado, envía una alarma que conduce a los análisis de investigación hacia el fraude, es decir, es una respuesta. Estos sistemas realizan operaciones conocidas como Análisis Absoluto de CDR's y son bastante eficientes para detectar los extremos de toda actividad fraudulenta. Sin embargo en los casos donde se desea realizar un Análisis Diferencial, se inicia haciendo monitorios a los patrones de comportamiento de los aparatos celulares, haciendo comparaciones de sus más recientes actividades, con la historia de uso del mismo. Tal como antes se ha mencionado, un cambio en el patrón de comportamiento es una de las característica sospechosas de la presencia de un escenario fraudulento (APEC, 1997)

En este tipo de análisis se pueden observar dos tipos de enfoques:

4.9 Marco conceptual

Dentro de este tipo de análisis, es necesario plantear casos observados en la práctica, que nos permitan hacer los análisis basados en casos prácticos, lo que nos permiten acercarnos más a la realidad. En el mejor de los casos pueden

ser usados las modalidades, indicando las causas de cómo se generan estos problemas y que sirva como herramienta que nos permita tomar decisiones ante estos casos, es decir, cuando nos toque ser víctima de este tipo de problemas. En los casos que se afronten problemas en un sistema basado en reglas, los casos prácticos antes mencionados, los estudios sobre comportamientos son estudiados en base a sus componentes de fraude; que luego se traducen a reglas compuestas por umbrales o medidas relativas. En el caso de las redes neuronales que se encuentran controladas y supervisadas, supervisadas donde se usan casos prácticos en forma de ejemplos de fraude y también los ejemplos de actos de usuarios sin relación con el fraude, para verificar mediante esta herramienta de control, cuáles son los comportamientos buenos y cuáles no lo son. Se considera que ambos tipos de herramientas deberían identificar comportamientos de alguna manera, en forma similar a los presentados en los actos prácticos, de esta forma en cuanto se presenten actos de fraude, produzcan alarmas, para superar el problema.

- **El aprendizaje**

Bajo este contexto, la herramienta usada será capaz de aprender el comportamiento veraz de un usuario legal que emitirá una alarma cuando dicho comportamiento haya cambiado sensiblemente. La habilidad de este tipo de herramientas para monitorear el movimiento de los usuarios que la hace sumamente útil para la detección de actos fraudulentos de los usuarios que no se dispone de ninguna información, así como todos los casos de fraude por suscripción, que se manifiestan en cambios de comportamiento. Si se dispone poco o nada acerca del fraude existente en el sistema, la técnica que se está proponiendo es una buena forma de operar para obtener buenas prácticas de comportamiento fraudulento; sin embargo, existen puntos analíticos importantes que debemos considerar, cuando se hace uso de este tipo de enfoques, entre los cuales se puede mencionar lo siguiente: No es posible enseñarle el uso de una herramienta, que buscar la misma, si los parámetros de evolución no se

configuran correctamente, porque se pueden cometer errores y no detectar cambios de comportamiento que lancen las alarmas correspondientes.

Visto en forma detallada este problema, es necesario plantear algunas alternativas de solución, pero se tendrá hacer en forma gradual, por lo cual se plantea lo siguiente:

- **Conocimientos sobre sobre Detección de cambios de comportamiento**

Para determinar si se han presentado o no cambios en el comportamiento de los usuarios, es preciso hacer comparaciones entre los perfiles CUP y UPH, midiendo la diferencia entre ambos, observando si es lo suficientemente grande como para lanzar una alarma. Se sabe que el CUP y el UPH son dos vectores que indican distribuciones de frecuencias, se puede emplear las distancias vectoriales para comparar qué tan, que tan diferentes pueden ser.

En este tipo de casos es bueno aplicar la distancia de Hellinger (H), en donde el valor que presenta muestras la diferencia entre sus distribuciones de frecuencias (Burge 2000). En cuanto a la distancia, siempre mantendrá un valor entre cero y dos; donde cero se emplea para distribuciones iguales y dos representa a la Ortogonalidad. En cuanto al valor de H, se indica que determinará el valor de las diferencias de las distribuciones de frecuencias CUP y UPH en el lanzamiento de alarmas, es decir cuando varía este valor, se activarán dichas alarmas. Tal como se ha visto antes, que través de los Registros detallados de llamadas y los Triggers, se pueden detectar las llamadas fraudulentas.

$$H = \sum_{i=1}^{\kappa} \sqrt{CUP_i - \sqrt{UPH_i}}$$

- **Alternativas de solución a la magnitud del problema**

La actitud que resaltada de las personas que generan este tipo de actos, como es el fraude de llamadas a celular, se indica que dependerá directamente

del **hardware** donde se registra las llamadas realizadas y bajo este sistema de control se puede detectar los actos dolosos o fraude, manifestados en el cambio de comportamiento en este sistema de control. Para ello se pueden considerar varios puntos; de acuerdo al software, la idea es mejorarla. En este trabajo operativo, es necesario usar lo menos posible con información relacional, dando mayor uso a archivos que contengan planos de datos, tratando en lo posible de usar informes escritos, sino en base a símbolos o lecturas de disco, que dependen de la comprensión. Como se puede observar esta posible solución, sólo hace uso de archivos y planos, almacenándose en un archivo por cada usuario con los datos distribuidos en CUP y UPH, así mismo todas las llamadas realizadas, hasta la última realizada.

- **Obstáculos a las propuestas de solución**

La solución a este tipo de problemas lo enfocamos en base a los problemas observados en base al cambio de comportamiento de los deferentes usuarios. Pero se tornaría tanto dificultoso detectarlo cuando el usuario realizara llamadas en cantidades altas, es decir , llamadas del mismo del tipo y además con niveles altos de consumo con, esto debido a que, haciendo comparaciones con su su patrón de comportamiento, siempre sería el mismo. Por lo cual para tener el mismo sistema de detección de fraude, es necesario hacer combinaciones de soluciones Otro modo operativo considerado como una limitación se puede observar en que los patrones de consumo se manifiestan en forma estática, situación que es diferente al consumo normal de los usuarios, para el mismo que es útil hacer reentrenamientos de las redes neuronales para así poder detectar la presencia de nuevos patrones de consumo, que figuren como intervalos o espacios nuevo patrón total de llamadas; y de esta forma considerar otros perfiles CUP y UPH, a partir de otras distribuciones.

V: METODOS Y PROCEDIMIENTOS

Métodos empleados

Se refiere a la metodología aplicada en este tipo de experimentos, lo cual se ha tenido que dividir en dos partes:

- 1° Se ha enfocado en la creación de patrones del usuario, las formas de construcción de los perfiles y el entrenamiento de las redes; y
- 2° En segundo lugar se enfocó el análisis de las llamadas de los usuarios posean altos niveles de consumo, considerando en este caso la detección de alarmas. A su vez, esta segunda prueba se ha tenido que dividir en dos partes:
 - a. La actualización de los patrones y así actualizar los perfiles UPH con la nominación ($f = 1$ llamada) y con el control de alarmas de cambio de comportamiento.
 - b. Se tiene que actualizar el perfil UPH diario una vez ($f = 1$ día).

5.1 Pruebas de creación de patrones de consumo

Para crear los patrones de llamadas locales, se han tenido que estructurar 3 redes neuronales. Se construyeron 3 redes neuronales, conocidos como Mapa de Auto Organización (MAO), para de esta manera crear los patrones de todos los tipos de llamadas, es decir: llamadas cortas o locales (LOC), DDN, o llamadas Nacionales (NAT) y DDI, llamadas a larga distancia o llamadas Internacionales (INT), en orden respectivo. Estas redes fueron puestas en prueba individualmente con una cantidad de llamadas referidas que realizaron durante cierto periodo de tiempo, controlado y administrado en los diferentes horarios. Las llamadas que se han indicado, además de ser las más comunes se ingresan a las redes en una forma, sin alterar el orden, pero sí en desorden alfabético, de forma tal, los patrones creados no fueron simplemente figurativos en los horarios y duración de las últimas llamadas. Los resultados que se obtuvo con esta experiencia logró definir en forma clara como un medio que permitía construir el perfil de los usuarios. Por lo general los patrones contienen la hora

y la duración en de las llamadas, contados en minutos; todo esto presentado de manera discreta.

. 5.2 Construcción de perfiles y detección de comportamientos

Si el investigador ya tiene determinado la cantidad necesaria de llamadas, se procede con la realización de la construcción de los perfiles de todos los usuarios a través de una distribución de frecuencias para cada uno de los patrones dentro de cada perfil (CUP y UPH) y la respectivas alarmas que se puede generar con la información empleada. El periodo que duró el proceso fue de tres meses usando como herramienta fundamental las llamadas durante ese lapso, tratando de que el reporte de ese consumo sea alto. En forma frecuente cada llamada actualizada dentro del perfil CUP del usuario, siempre es comparaba con las del perfil UPH, de esta manera encontrando entre la distancia Hellinger (H) entre ambos. Pero es importante tener presente que si esta respuesta supera a la dimensión fijada como límite, En función a la distribución de las frecuencias se actualiza el comportamiento del perfil UPH con el apoyo del CUP, según se comporten los elementos. Es importante comentar, que los procesos de construcción y actualización, siempre se ha realizado desde el preciso momento en que se produce la primera llamada del usuario; sin embargo, y como es normal, la comparación y su correspondiente respuesta de detección de señales se han obtenido luego de hacer comparaciones y análisis de toda la información, es decir, después de procesar la cantidad de llamadas analizadas para que el usuario pasara la cantidad mínima para construir un perfil (QL). Con la necesaria información referida al usuario. En forma descriptiva se puede sostener que al momento de observarse el ingreso de la primera llamada del usuario, todos los patrones del CUP y UPH tienen movimientos constantes y muy sensibles originado por la distribución de frecuencia, suponiendo que el usuario tiene tendencias promedias para realizar cualquier tipo de llamada, sin información. Como mínimo este tipo de experiencias se realizan dos veces consecutivas: la primera vez es para actualizar el UPH, por cada llamada realizada y por consiguiente se observa un bajo umbral Hellinger (H, como medida promedia) para la detección de fraude a través de alarmas, esto debido a la diferencia, con posibilidad que se presente entre los perfiles CUP y UPH, pero estas posibilidades son bastante pequeñas, logrados con la actualizando

el perfil histórico de cada llamada. Esta misma forma de operar se podría considerar de igual manera para el perfil actual. La segunda experimentación que mencionaremos se desarrolló s actualizando el UPH, una vez por día y un umbral Hellinger (H) lo suficientemente alto para observar las diferencias que puedan manifestar cambios de comportamiento.

5.3 Criterios aplicados en la creación de patrones informativos

Se resume en los siguientes criterios:

- Tamaño de la red neuronal que tenga la capacidad de clasificar todas las llamadas locales, es decir $NL \times ML = 12 \times 12$.
- Tamaño de la red neuronal con una capacidad de clasificar todas las llamadas nacionales, es decir $NN \times MN = 8 \times 8$
- Tamaña de la red neuronal con la capacidad poder clasificar todas las llamadas internacionales, $NI \times MI = 6 \times 6$
- El aprendizaje estático (α) = 0,6 %
- También es importante considerar la Ubicación de neurona “vecina” que tuviera relación ($DVMAX$) = 10

Estos criterios aplicados indican la dimensión de los perfiles creados CUP y UPH

- Diferentes patrones para ordenar las llamadas LOC (PL) = 144
- Patrones para ordenar llamadas NAT- DDN (PN) = 64
- Registros varios para ordenar las llamadas INT (PI) = 36
- Tamaño total de los perfiles CUP y UPH (K) = 244

5.4. Tamaño de la estructuración de perfiles: alteración de comportamiento del usuario

Los métodos empleados en la elaboración de perfiles y detección de alteración, se consideraron los siguientes:

Prueba N° 1:

- El problema que se observa en la adaptación de las llamadas LOC en el CUP (a LOC) = 0,8
- El problema que adaptación de llamadas LOC en el CUP (a NAT) = 0,8
- El problema que adaptan de llamadas locales en el CUP (a INT) = 0,8
- El problema de adaptación del UPH (b) = 0,9
- La capacidad de determinación de grado de sensibilidad del sistema,- Umbral Hellinger (H) = 0,3
- Vistas las acciones con qué frecuencia se actualizan los UPH (f) = 1 llamada.

Prueba N° 2

- La adaptación de llamadas LOC en el CUP (a LOC) = 0,8
- La adaptación de llamadas LOC en el CUP (a NAT) = 0,9
- La adaptación de llamadas LOC en el CUP (a INT) = 0,9
- La adaptación del UPH (b) = 0,6
- El grado de sensibilidad del sistema ante el problema - Umbral Hellinger (H) = 0,75
- La frecuencia con que se actualiza el UPH, tiempo mínimo un día (f = 1 día).

5.5 Valores asignadas a las pruebas:

En la comparación de los perfiles se requiere como mínimo 100 llamadas, a la cual se le puede representar por (QL).

Complementariamente a las alternativas planteadas en los casos de experimentación, al respecto es importante ampliar los argumentos de la siguiente manera:

En la primera parte presentamos la información requerida, lo cual nos ha permitido realizar la construcción de los perfiles antes indicados y el entrenamiento de las redes neuronales. En esta parte del trabajo se observó la influencia de las variables en forma determinante; es decir, es una descripción de los experimentos, a la espera de los resultados que se plantearán después.

Información empleada en la experimentación

- Se deben clasificar todas las llamadas para que las redes neuronales tengan facilidad en su procesamiento.
- Las dimensiones consideradas para clasificar las llamadas son:
 - a.- Tipo de llamadas, que define a cuál de las redes neuronales corresponde, es decir, si son LOC, NAT o INT.
 - b. Hora de la llamada, Se tiene que analizar en la franja horaria donde se realizó la llamada para ver los intervalos, por ejemplo, si la llamada se produjo a las 23.15 horas, corresponde a la franja n23; si se realizó a las 8.32, corresponde a la franja 8. Para ser procesado en las redes SOM (Mapa Auto Organizado, se debe normalizar a un intervalo que oscile entre 0 y 1

Generalmente el tiempo de duración de cada llamada la duración de cada llamada a tiempos completos que pueden ser minutos completos y nada de fracciones inferiores para que las redes SOM hagan un buen trabajo de procesamiento, tal como antes se indicó que el Mapa Auto Organizado esté dentro de un intervalo entre 0 y 1; además se debe tener en cuenta que el tiempo máximo de llamadas deben ser 30 Min.; esto indica que las llamadas mayores a este tiempo considerado como máximo, no influyen en el comportamiento de las redes SOM.

En resumen las dimensiones que tome cada neurona de las redes consideradas contienen dos valores: **Hora de la llamada valorada en el intervalo entre 0 y 1; duración de la llamada valorada en el intervalo entre 0 y 1.**

Las dimensiones o tamaños empleados para la clasificación de una llamada son:

- **Reconocimiento de la llamada:** Es decir, es importante reconocer, de qué tipo de llamada se trata: LOC (Llamada local); NAT (Llamadas con destino Nacional) y/o INT (Llamadas con destino Internacional).
- **La hora en que se realiza la llamada:** Es importante determinar en qué franja y qué hora es la que se realiza la llamada, lo cual está distribuido entre 24 intervalos. El análisis de los intervalos con respecto a este; solamente se recuerda que para procesar la información en las redes SOM, siempre se normaliza el valor en un intervalo entre 0 y 1.
- **El tiempo que dura cada llamada:** Para facilitar las llamadas, es necesario reducir todas las llamadas a minutos; y de esta manera facilitar el procesamiento en las redes del Mapa Auto Organizado. Todas las llamadas se redondean a minutos completos para ser procesados por las redes del Mapa Auto Organizado (SOM), se ha normalizado el valor en un intervalo entre 0 y 1, considerando como valor máximo: 30 minutos, siempre recordando que todas las llamadas por celular no deben exceder más de 30 minutos; si existe llamadas superiores a este tiempo, se pueden considerar como si fuera otra llamada.

5.6 LAS REDES NEURONALES: PRUEBAS

Para la pruebas de acción y funcionamiento de las redes neuronales que han definido que han definido los patrones, empleado en elaboración de los perfiles del usuario, considerándose en estos casos las llamadas reales de más de un día realizados por todos los usuarios de la red; acción que permitió determinar que los patrones de uso que formaran partes de las redes, adquieran información de todo el espacio de tipo llamadas diferentes.

Para prueba si las redes neuronales para las llamadas locales se emplean 650,000 llamadas sin detallar orden alguno, en este caso se entiende que se ha “clasificado” las llamadas en patrones similares a las últimas llamadas procesadas; siempre que estas fueran presentadas al azar. En la prueba de la red neuronal correspondiente a las llamadas NAT, la prueba se realizó sobre 65.000 llamadas; en la primera parte de la prueba de la red neuronal de

llamadas internacionales se utilizaron 2.100 llamadas. El formato de registro empleado fue el siguiente:

**(AAAAMMDD) + (HH24MISS) + (DURACION en FORMATO 00000) +
(TIPO DE LLAMADA – LOC, NAT o INT)**

- Perfiles del usuario y Detección de cambios de comportamiento

Si se determina la imposibilidad de trabajar con casos específicos y verificación de cambios de comportamientos de los usuarios, casos donde las empresas no disponen nada de antecedentes de los clientes con respecto a cambios fraudulentos, se toman decisiones junto a los técnicos de analizar los campos de alto nivel de consumo de llamadas Nacionales e Internacionales, cuyos comportamientos si manifiestan duda. Si las pruebas mostraban sospechas en el comportamiento de los usuarios, se estaría demostrando la presencia de una tendencia de cambio de comportamiento doloso, donde el tiempo de uso no coincide con los costos generados, indicando para la Empresa ordenaría análisis de manera especial con el fin de determinar el fraude.

Asimismo se procesaron todas las llamadas de los usuarios desde enero 2003 a marzo 2016 en un promedio de 100,000 llamadas, con las cuales se han construido los perfiles de cada uno de los CUP y UPH, llegando a detectar las alarmas la producción de alarmas. Todas estas situaciones merecieron considerarlo como un indicio de fraude.

Una vez logradas la presencia de las alarmas las variaciones de comportamiento, las llamadas deben ser analizadas de una forma detallada, en lo que respecta a cada usuario, para determinar si en verdad hubo cambios de comportamiento las llamadas de cada usuario y determinar si en verdad, si hubo cambios en su comportamiento, Para la elaboración de los perfiles del usuario es bastante importante conocer y relacionar cuantos patrones de uso se pueden

utilizar, para de esta manera trabajar en forma discreta en el campo de llamadas para la detección de fraude, considerando la sensibilidad del sistema

HIPÓTESIS Y VARIABLES

Planteamiento de Hipótesis

Ha. Hipótesis General

Se ha determinado que existe una relación significativa entre la detección de fraude en el uso de telefonía celular con el uso de las Redes Neuronales en Lima distrito, periodo 2015.

Hipótesis Específicas

H₁. Se ha podido establecer un alto grado de relación entre el fraude en el uso de telefonía celular con los costos de los servicios de comunicación en Lima distrito, año 2015.

H₂. Existe una relación bastante considerable entre la detección de fraude en el uso de la telefonía celular y las estructuras de patrones de llamada LOC, NAY y INT en Lima distrito, año 2015.

H₃. Los estudios demuestran que existe una relación determinante entre el fraude en el uso y control con las redes neuronales y la construcción y mantenimiento de los perfiles del usuario en Lima distrito, periodo 2015.

- VARIABLES

a) Independientes

- Las redes neuronales en los procesos donde se busca determinar en forma clasificada los diferentes tipos de llamadas, siempre requieren de dimensiones lo suficientemente grande como para poder hacer clasificaciones amplias. En cuanto a las llamadas locales indica una dimensión NL x ML, que describe el Mapa de Auto Organización (SOM), que permite clasificar las llamadas. Locales.

- Disponer de la Red Neuronal con una dimensión, útil para determinar la clasificación de las llamadas Nacionales (NNx MN) , La cual tiene una dimensión de la matriz que hace descripciones del Mapa de Auto Organización (SOM) para clasificar las llamadas Nacionales .
- El tamaño de la red neuronal, Útil para clasificar todas las llamadas Internacionales (NI x MI), la misma que define el desarrollo de la, matriz que permite describir el Mapa de Auto Organización (SOM), en el caso del análisis de las llamadas internacionales.
- Con este mismo trato las tasas de aprendizaje estarían presentando un comportamiento estático (α): Es la misma tasa estática de aprendizaje que nos permite adaptar cada una de las neuronas (patrones) de nuestra red. El valor de α se tiene que determinar técnica experimentalmente en valores más adecuados, que oscilan entre 0,6 y 0,9. Si estos valores fueran más inferiores, demostrando la gran sensibilidad de cada nueva llamada, donde se observará los cambios en los patrones presentado en forma rápida actitud por el cual las llamadas anteriores tienden a perderse, en cierta forma distorsionando así los estudios realizados.
- Es importante definir la distancia máxima que se debe considerar entre neurona vecina con una neurona afectada (DVMAX). Los análisis en estos casos nos permiten los patrones actualizados en el entrenamiento de las redes SOM. Lo importante es hacer las definiciones y mediciones de las funciones de las neuronas vecinas, que en nuestros análisis se han convertido en dependientes de la variable DVMAX.

Es necesario recordar las reglas de aprendizaje en un SOM (Hollmen, 1996):

$$M_i(t_1) = m_i(t) + h_{ci}(t)(x(t) - m_i(t))$$

Donde:

$m_i(t_1)$ = Representa al peso de la neurona i en el tiempo t_1

$m_i(t)$ = Representa al peso de la neurona i en el tiempo t.

$x(t)$ = Representa al patrón de entrenamiento t (o llamada codificada).

$H_{ci}(t)$ = Representa a la función de vecindario en el tiempo t , definida como:

$$h_{ti}(g) = \alpha(g) \frac{D_{VMAX} - d(t)}{D_{VMAX}}$$

Donde:

$\alpha(t) = \alpha(1-g/gMAX)$ (con α : tasa de aprendizaje estática): tasa de aprendizaje en el tiempo g .

$d(g)$ = Distancia entre la neurona hasta la neurona vecina, es decir, a la neurona BMU en el tiempo g (A medida que el tiempo g avanza, $h_{ci}(g)$ tiende a cero; lo que indica, que el vecindario se hace más pequeño y las neuronas vecinas que se actualizan disminuyen).

b) Variables dependientes

a) La lista de patrones de uso empleados para la clasificación de las llamadas locales (PL): indica que estos patrones se utilizarán en la red neuronal para las llamadas locales, para de esta manera, el mismo que se obtiene multiplicando el NL x ML.

b) El rol de patrones necesarios para la clasificación de las llamadas DDN (PN), en lo referente al listado que se empleará en el control de la red neuronal de diferentes tipos de llamadas, para analizar todo el espacio de llamadas; valor que se calcula multiplicando NN x MN, es decir en lo referente a las llamadas nacionales.

c) En lo que respecta a los listados utilizados en la clasificación de las llamadas con destino Internacional (PI), relacionado con la cantidad de patrones de los usuarios que emplearemos en la red neuronal para las todas las llamadas

Internacionales para clasificar todo el espacio de llamadas en un tiempo de análisis previamente establecido, obteniéndose al operar $NI \times MI$,

d) En que respecta a las dimensiones de los perfiles CUP y UPH (K): En los cuales se indican la cantidad total de patrones, que contienen los perfiles de los PL patrones LOC, PN patrones NAT y PI patrones INT.

- **Funcionamiento de cada variable**

a) Independientes

- El factor principal en la adaptación de llamadas locales en el CUP (α_{LOC}), el mismo que indica el % de adaptación cuando es aplicada a una llamada X, tiende a incorporarse al CUP, siempre y cuando la llamada corresponda a una llamada local.

$$CUP_i = \alpha_{Loc} CUP_i - (1 - \alpha_{Loc}) X_i, \text{ con } 1 \leq i \leq k$$

- El elemento de adaptación del control de llamadas locales en el CUP (α_{NAT}), referido a la % de adaptación detectada cuando una llamada cualquiera que sea, sea incorporada al CUP, siempre que esta llamada corresponde a una llamada Internacional.

$$CUP_i = \alpha_{NAT} CUP_i - (1 - \alpha_{NAT}) X_i, \text{ cuando } 1 \leq i \leq k$$

- El elemento que se adapta al sistema de llamadas locales dentro del CUP (α_{INT}), que representa a la % de adaptación que se aplica cuando una llamada cualquiera que sea, se incorpora al CUP, siempre que esta llamada corresponda a una llamada internacional (DDI).

$$\mathbf{CUP_i = \alpha_{INT} CUP_{i-1} + (1 - \alpha_{INT})X_i, \text{ con } 1 \leq i \leq k}$$

- El elemento de adaptación del UPH (β), que se corresponde a la % de adaptación que se aplica cuando el CUP, siempre que este instrumento se incorpore al UPH.

$$\mathbf{UPH_i = \beta UPH_i + (1 - \beta) CUP_i, \text{ cuando } 1 \leq i \leq k}$$

- En el análisis del contenido de esta ecuación es necesario ver que la frecuencia de actualización del UPH (f), el mismo que está definido en el intervalo de tiempo en el que se actualizará el perfil UPH, supuestamente dentro de cada llamada en n horas, cada cierto número de días. es decir, dependiendo del tiempo que se programe o se estime conveniente, el mismo que se ajusta además de la frecuencia del movimiento de las llamadas.

- Las cantidades mínimas de llamadas antes de hacer las comparaciones de los perfiles (QL), es necesario de realizar ningún tipo de comparaciones los perfiles CUP y UPH con el fin de determinar si hubo o no cambios de comportamiento, según la información relevante, poseer los perfiles de los usuarios con la información oportuna, que haga descripción correcta del patrón del comportamiento del cliente; es decir, que las QL llamadas iniciales del usuario sirvan para construir su perfil, donde no es tan necesario buscar las QL+1 Alarmas).

b) Dependientes

- Como es natural, la sensibilidad con respecto a cualquier problema, siempre se asienta en las variables dependientes, debido a que estas son o sufren las consecuencias de las variables causas, por esta razón es que dentro del sistema se le considera como "Umbral", según Hellinger (H): quien lo consideraba como,

Helliger lo definía como una medida entre dos distancias del tipo de vector y su valor indica la diferencia entre dos distribuciones de frecuencias. Burge & Taylor, en 1997, al respecto decían: que la distancia siempre es y será un valor entre cero y dos; cero siempre representa a las distribuciones iguales y dos representa a la ortogonalidad. El valor asignado a H indica la diferencia de la distribución de frecuencia de CUP y UPH para lanzamiento de alarmas. Según los autores, este valor generalmente se mide mediante la siguiente ecuación:

$$H = \sum_{i=t}^k (\sqrt{CUP_i} - \sqrt{UPH_i})^2$$

Luego de haber determinado todos estos valores que se está mencionando y visto la experimentación respectiva, es preciso analizar de uno en uno las alarmas que se han lanzado con el sistema, siempre considerando el detalle de llamadas de cada usuario, observando si realmente se han generado cambios significativos en el comportamiento de los clientes.

Operacionalización de las Variables

Operacionalización de la Variable 1: Detección de fraude en telefonía Celular

Operacionalización del fraude en telefonía celular

DIMENSIONES	INDICADORES	ITEMS	ESCALA	NIVEL- RANGO
Costos de los servicios de telefonía	-Impago de facturas como efecto del fraude	1, 2, 3,	NUNCA	POSITIVO 101 – 135
	-Mayor tecnología de los servicios.	4, 5, 6,		
	-Satisfacción de los usuarios	7, 8, 9		

Estructuras de patrones de llamada LOC, NAT y INT.	-Facilidad en el control de los usuarios	10, 11, 12,	CASI NUNCA	MODERADO 64 – 100
	-Diferencia de costos por los servicios	13,14, 15,		
	-Medición de la rentabilidad entre los tipos de llamadas.	16,17, 18,		
Construcción y mantenimiento de los perfiles del usuario	-La satisfacción generada en el usuario.	19,20, 21,	CASI SIEMPRE	NEGATIVO 27 – 63
	-Los costos de mantener bien los perfiles.	22, 23, 24,		
	-Los beneficios obtenidos.	25, 26, 27	SIEMPRE	

FUENTE: Elaboración propia del investigador

Tabla N°01

Las variables consideradas en la detección del fraude en las llamadas de telefonía celular se han tomado en forma técnica, en base al tipo de problema, relacionándolo las variables independientes con las dependientes, de tal manera que tales variables manifiesten una verdadera dimensión del problema de la sociedad perjudicada y la manera cómo se manifiestan sus efectos es la sociedad de usuarios y la empresa, es decir sus indicadores, de qué forma lo observamos para convertirse en un problema social

a) Operacionalización de la variable 2: Utilización de las Redes Neuronales

Tabla No. 2

Operacionalización de la Utilización de las Redes Neuronales

DIMENSIONES	INDICADORES	ITEMS	ESCALA	NIVEL- RANGO
El robo de línea como elemento característico del fraude	-Las técnicas empleadas en el robo de línea	1, 2, 3,	NUNCA	POSITIVO 101 – 135
	-Los daños cometidos en los usuarios legales.	4, 5, 6,		
	-La frecuencia con que se comete el robo.	7, 8, 9		

El comportamiento de los afectados por el robo de línea	-Los reclamos sin solución al problema.	10,11,12,	CASI NUNCA A VECES	MODERADO 64 – 100
	-Los montos elevados que están fuera del de los presupuestos	13,14, 15,		
	-El abandono al uso del servicio.	16,17, 18,		
El efecto social que genera este problema.	-Pérdida de ingresos del fisco por impago de impuestos.	19, 20, 21,	CASI SIEMPRE SIEMPRE	NEGATIVO 27 – 63
	-El incremento del fraude por esta modalidad.	22, 23, 24,		
	-El descontrol por parte de las instituciones competentes	25, 26, 27		

FUENTE: Creatividad del investigador

Tabla N° 02

En la tabla N°02 se observa como los individuos que incurren en este tipo de falta consideradas como delitos, hacen uso de línea telefónica sin realizar ningún pago por este servicio, generando problemas para la empresa, para los usuarios propietarios de la línea y para el Estado, es decir, los efectos del dolo, tal como se expresa continuación:

En cuanto al primer caso, en cuanto se refiere al robo de línea como elemento característico de esta acción, manifestándose a través de las técnicas de operación empleadas en el dolo, acciones que lejos de disminuirlas, se incrementan más, generando daños diversos en los usuarios, que finalmente tienen que abandonar el servicio, haciendo un gran perjuicio a la empresa, resultando aún más dañino por la frecuencia con que lo hacen.

Todas estas acciones hacen que los clientes legales de las empresas de este tipo de servicios, en principio reclamen a las empresas proveedoras, reclamos sin ninguna solución, ante esta negativa tomar la decisión de abandonar el servicio y la empresa exigir al pago, usando medios legales, hasta afectarles con imposibilidad de créditos financieros.

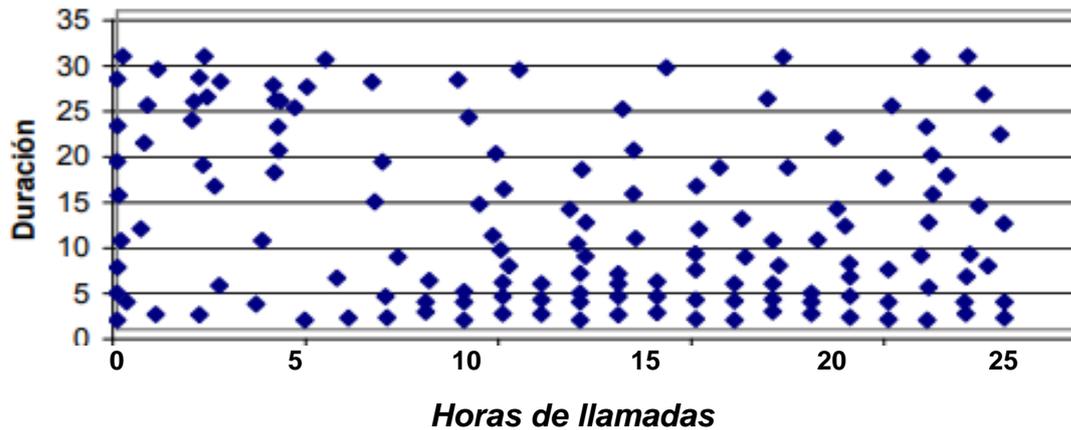
En el aspecto social se observa que las empresas proveedoras del servicio, los impuestos generados por este medio tienden a evadir este tipo de pagos, generando un gran problema para fisco. Asimismo se observa que este tipo de acciones fraudulentas tienen tendencias crecientes, por lo rentable que resulta para los individuos beneficiados; y ante el crecimiento doloso, las entidades de control, se sienten incompetentes de controlar, además, no es de sus competencias velar por este problema, sino de exigir que las empresas vendedoras del servicio que cumplan con los usuarios legales y que éstas cumplan con sus obligaciones con el Estado.

- **Con respecto a la creación de patrones**

Como técnicamente es usual arribar a esta parte de la investigación, presentamos las pruebas de los resultados del trabajo obtenidos luego de aplicar las pruebas correspondientes a las 3 redes neuronales utilizadas en los análisis; es decir, que estos resultados muestran, los efectos de cada uno de los patrones de uso que las redes lograron determinar, teniendo en cuenta que sean los más representativos del espacio de todas las llamadas que realizan los usuarios. A continuación presentamos cada uno de los casos:

- a) Estamos presentando tres gráficos en donde se observan los listados o patrones creados; en el eje de las abscisas se puede ver la hora en que se produjo las llamadas, mientras que en el eje de las ordenadas presentamos la duración de tales llamadas y los puntos dentro del gráfico indican cada uno de los patrones elegidos por la red. El gráfico N° correspondiente a la red neuronal que estudia las llamadas locales, representando a 144 patrones; mientras que en lo que corresponde a llamadas nacionales, la red neuronal DDN, necesariamente considera 64 patrones; y en el de la red DDI que corresponde a las llamadas Internacionales, se observan 36 patrones..

Patrones de llamadas locales

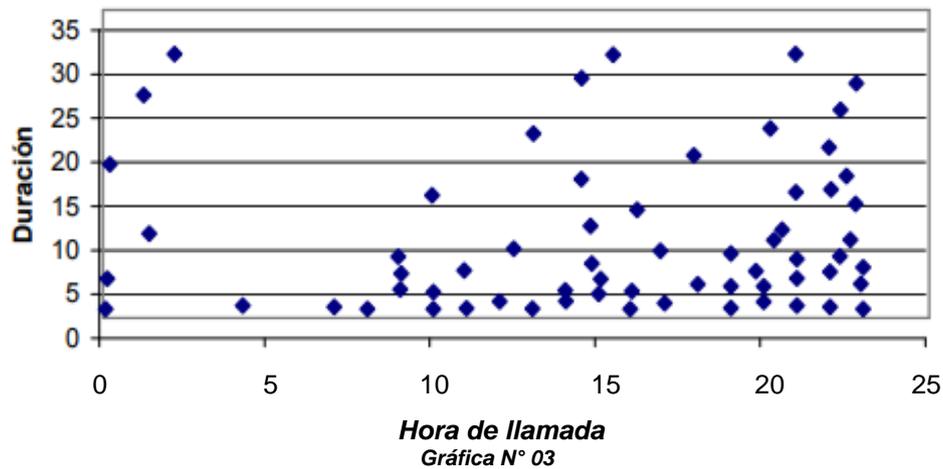


Gráfica N° 02

Las 144 puntuaciones que corresponden a los patrones de llamadas nacionales, después de la prueba realizada con la red neuronal observándolo en forma rápida se nota que existe una gran concentración de patrones en el intervalo de las 08.00 horas hasta las 20.00 horas, con una duración promedio de 0 a 5 minutos; indicando que todas las llamadas locales generados por esta empresa se realizan dentro de la hora y tiempo indicados en este registro.

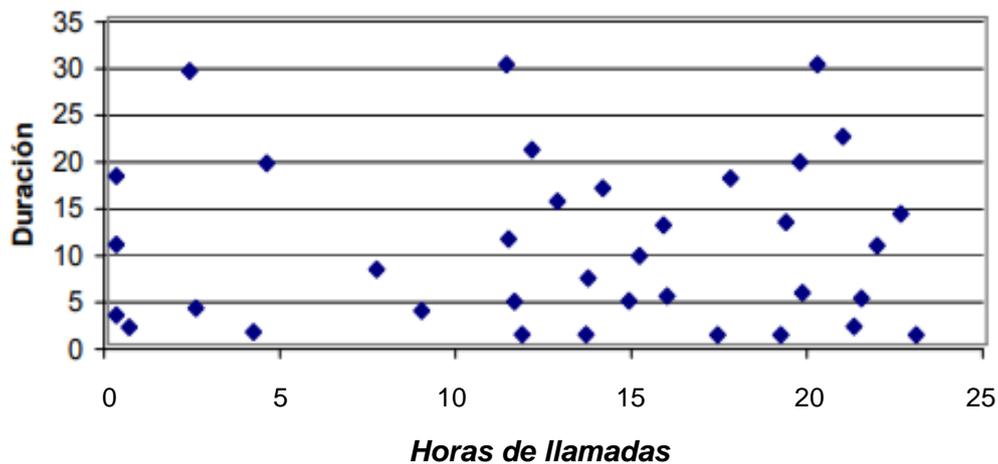
- b) En segundo término se realizó la prueba de las llamadas nacionales Siempre siguiendo el método de análisis de las llamadas locales, se observa la presencia de 64 patrones creados luego de la prueba de la red neuronal de llamadas nacionales, tal como se puede ver la puntuación en el gráfico N| 03

Patrones de llamadas nacionales DDN



En la gráfica N°03 se observa una concentración de llamadas en los horarios entre el intervalo de 15 a las 22 horas, con duraciones que oscilan entre 0 minutos hasta 7 minutos en promedio. Esto indica que hay patrones creados para horas de la madrugada, concluyéndose que la mayoría de los usuarios - clientes de la empresa en cuestión por lo general no hacen uso de las redes DDN en horas de movimiento masivo, para evitar los cruces de llamadas, es decir, aprovechan la mayor nitidez de la línea.

- c) En el tercer caso, siguiendo en forma secuencial el mismo tipo de análisis, en lo referente al uso de las llamadas internacionales y de acuerdo a la información obtenida y empleada, se puede ver que existen 36 patrones de uso creados luego de las pruebas realizadas sobre las llamadas internacionales. La distribución de la información, por lo general se realiza en forma aleatoria, pero la duración de las llamadas consideradas como patrones tienen un promedio de duración entre 7 y 10 minutos de duración, tal información lo podemos comprobar en la gráfica N°04.



Gráfica N° 04

Luego presentamos a manera de información la matriz U-MATRIX que muestran la idea, de cómo se han formen las zonas que manifiestan distintas categorías entre los grupos de llamadas.

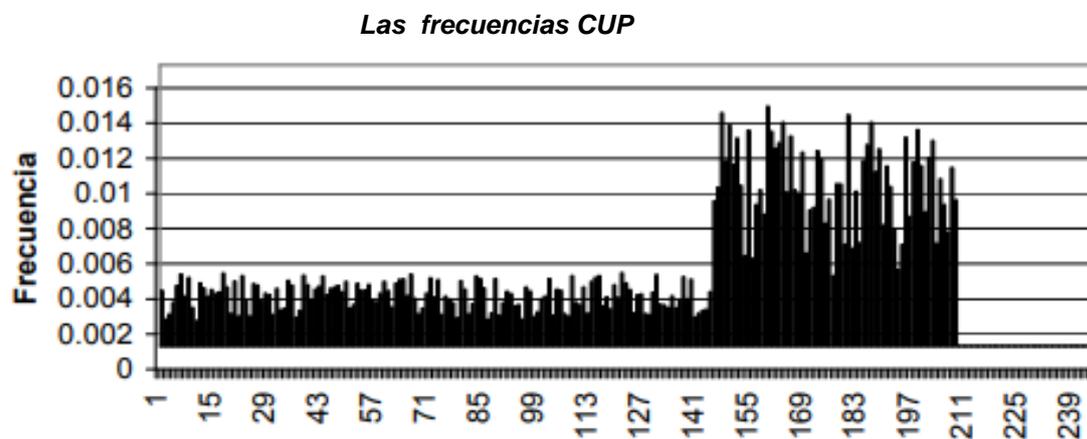
- Los perfiles como medios de cambio de comportamiento

Es bastante importante presentar los resultados de la investigación obtenidos después de haber presentado los perfiles debidamente estructurados para determinar la señal de las alarmas para cada experiencia realizada la construcción de los perfiles y los medios de detección de las correspondientes alarmas para cada una de las experiencias realizadas. Tal como hemos podido ver en las gráficas presentadas la descripción de los perfiles CUP y UPH en los diferentes temas, en los instantes precisos en que se lanzó la alarma.

En el eje que corresponde a las X existen 244 patrones, distribuidos de la siguiente manera: 144 para llamadas LOC, 64 para llamadas NAT y 36 para llamadas INT. En el otro eje Y presentamos las frecuencias de uso de cada uno de los usuarios o listados disponibles para cada usuario, observado en el preciso instante en que la alarma fue lanzada. También tendrá lugar una explicación general con respecto al porqué el sistema puede generar alarmas y la confiabilidad mostrada, hasta que punto muestran credibilidad, considerando que todo está generado por el detalle dado en base a las llamadas de cada usuario.

El UPH actualizado con cada llamada

Prueba N° 1



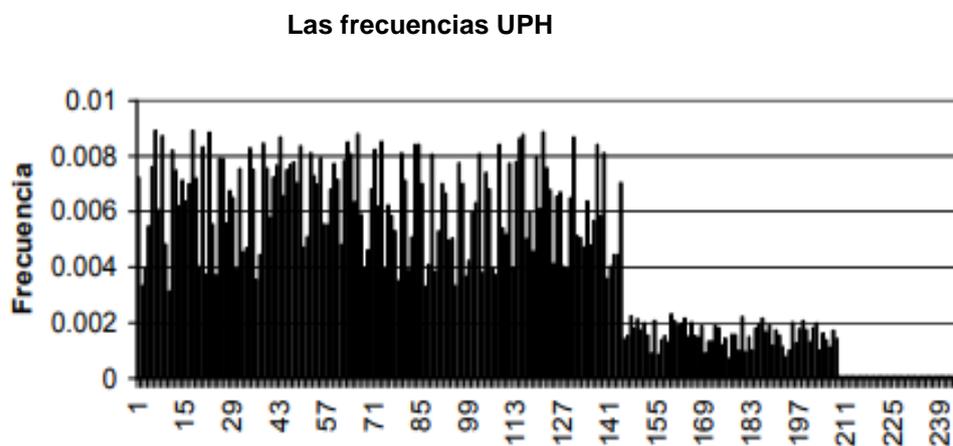
Patrones

(LOC: 1 – 144; NAT.: 145 – N 208; INT.: 209 – 244)

Gráfica N° 05

En el gráfico N° 05 que corresponde a la muestra de las pruebas CUP de un usuario cualquiera en el momento en que se produce un lanzamiento de alarma. En este lanzamiento se puede observar que la distribución de frecuencias presenta una tendencia creciente de llamadas nacionales (DDN) de 145 a 208.

Prueba N° 02



Patrones

(LOC: 1 – 144; NAT: 145 – 208; INT: 209 – 244)

Gráfica N° 06

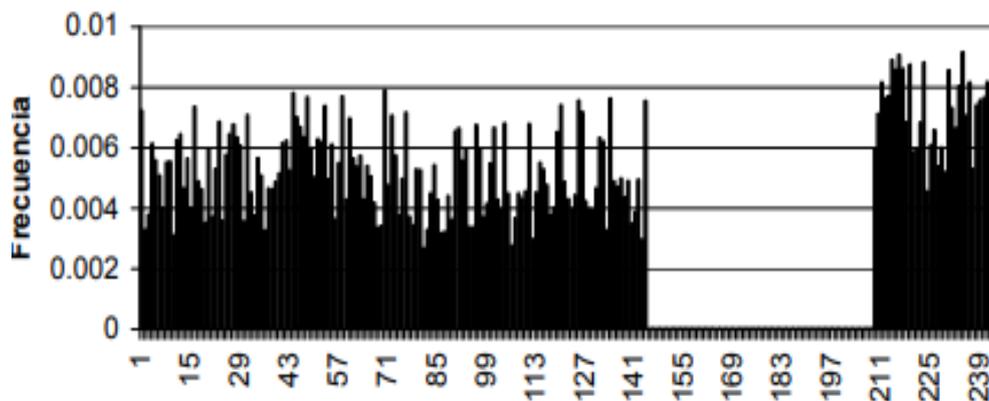
El gráfico N° 06 nos muestra el UPH del mismo usuario en el momento preciso en que la alarma ha sido se lanzada. En esto se puede notar que las frecuencias distribuidas en el eje, indica en sus resultados una tendencia más alta para la ejecución de llamadas dentro de los patrones de 1 a 144.

Se puede comentar como resultado que la diferencia entre las distribuciones de frecuencias, sostenida por Hellinger (H) es igual, a: 0,30081. Haciendo una ampliación del usuario de las llamadas de dicho usuario, considerando los antecedentes de sus llamadas al lanzamiento de la alarma hasta la aparición de la misma, se puede observar que la alarma se activó, como efecto a que el usuario efectuó una llamada DDN por vez primera desde que sus llamadas fueron procesadas. Esto indica que su patrón de comportamiento tenía que variar por la ejecución de llamadas de este tipo histórico no precisaba que iba a ejecutar llamadas de este tipo. En efecto al ser analizadas, el instrumento de investigación aplicado detectó cambio, activan luego la alarma. Además se presentan, los cuales llevados a la prueba, presentando alta sensibilidad, por más diferente que sea, puede generar un cambio de comportamiento que activa a la alarma. Todas las alarmas que han sido lanzadas, luego de ser analizadas, los 60 usuarios observados de un total de 88, de los cuales 33 corresponden a diferentes situaciones. Todo ello se debe, una vez que se ha lanzado una alarma a la llamada de un usuario, las llamadas siguientes del mismo lanzan nuevamente alarmas, a tal punto que el UPH tiende a no adaptarse en forma definitiva al cambio de comportamiento. A su vez indica que en su mayoría siguen el mismo patrón de cambio, considerándolo como normal el cambio de comportamiento; de esta manera se logra determinar como que el usuario como sospechoso de fraude, materia de la presente investigación. .

Sensibilidad del UPH analizado con el Umbral Helliger (H)

Prueba N° 02

Orden de las frecuencias CUP



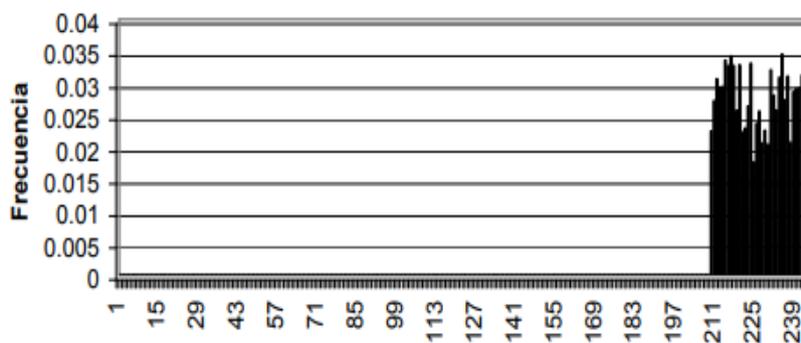
Patrones

(LOC: 1 – 144; NAT: 145 – 208; INT: 209 – 244)

Gráfica N° 07

En la gráfica N° 07 se muestra las pruebas de un usuario en el preciso instante en que fue lanzada una alarma respondiendo a su comportamiento. Así mismo se puede observar el orden de las frecuencias, cómo fueron distribuidas, manifestando la tendencia de las llamadas locales en base a los patrones de 1 a 144 e internacional en los patrones del 209 a 244.

Distribución de frecuencias UPH



Patrones

(LOC: 1 – 144; NAT: 145 – 208; INT: 209 – 244)

Grafica N° 08

En la gráfica N° 08, observamos el UPH del usuario en mención en el preciso instante en que la alarma es lanzada. En este punto es fácil distinguir, cómo las frecuencias se distribuyen dentro del eje respectivo, en lo que respecta a las llamadas internacionales, considerando los patrones que van desde 209 a 244. Por lo tanto, la diferencia observada en la distribución de sus frecuencias analizada por la distancia Hellinger (H) es 0.82815. Haciendo un análisis en forma detallada con respecto a las llamadas de este usuario, considerando su comportamiento histórico hasta el lanzamiento de la alarma, indica que la activación de la alarma se produjo como efecto en que el usuario solamente realizaba llamadas internacionales, pero en cierto momento no determinado empezó ejecutar llamadas locales. Pero cuando la cantidad de llamadas locales generó un cambio en el CUP de tal forma como se muestra gráficamente, la alarma. En la práctica, resulta un acto bastante curioso, que corre el riesgo que la acción de la alarma no fuera un buen indicador de fraude, si el usuario hace sus pagos de facturas de llamadas internacionales. Pero analizándolo técnicamente, se refiere a un indicador, analizado el cambio a través de la sensibilidad, mostrándose el efecto en el cambio de comportamiento en su patrón de consumo y esto lo que busca el sistema. Todas las alarmas que han sido lanzadas, siempre los análisis se han realizado sobre 60 usuarios, de los 64 disponibles; de las cuales 14 corresponden a otros casos. Generalmente esto se debe a que luego de haber sido lanzada la alarma de un usuario para este tipo de casos, por la relación que existe entre ellas y la sensibilidad que poseen, es bastante probable a que también se activen las demás; lo que indica también que el UPH no se adapta definitivamente al cambio de comportamiento del usuario. Este tipo de cambios tienden a acentuarse más en el UPH, debido a que en un inicio cuando se procesa la información o la llamada del día anterior, se actualiza. La su gran mayoría en este tipo de cambios, seguidos del patrón de gráficos el mismo que debe contener varias llamadas fuera del patrón de comportamiento para que el sistema pueda encontrar al usuario sospechoso. Estos resultados obtenidos son bastante satisfactorios dentro de la prueba N° 01, donde el usuario manifiesta bastante sensibilidad, generando sospecha de fraude en su consumo, o también la sospecha puede darse por el simple hecho de ejecutar una llamada diferente.

ESTRATEGIAS METODOLÓGICAS

5.8. Tipo de investigación

De acuerdo a los análisis realizados en este trabajo, este estudio se encuentra dentro de los trabajos para ser considerado como una investigación teórica experimental, por las razones a que en su desarrollo se han empleado informaciones propias de los procesos administrativos, orientados a los procesos de gestión y de conducción de las empresas que proveen servicios.

Nivel de investigación

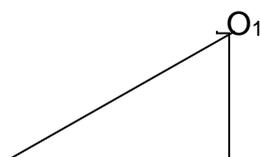
En función a las naturales seguidas en proceso de este trabajo, se observa que reúne las condiciones para ser Descriptivo, Explicativo y Correlacional.

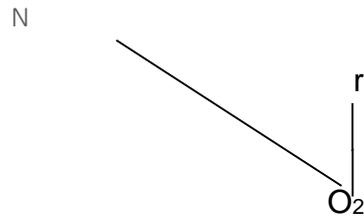
Diseño

El presente trabajo de investigación está basado en el plan general del investigador, es decir, a las decisiones tomadas y en función a las respuestas a sus interrogantes o al comprobar las hipótesis de la investigación. Tal como se ha mencionado antes, al este estudio del tipo Descriptivo Correlacional, por que busca relacionar las variables Independientes y dependientes para medir la magnitud del problema en Lima distrito, en el periodo 2015, miden el grado de asociación entre dos o más variables (cuantifican relaciones), es decir, miden el estado y la dimensión.

También miden y analizan, es decir, lo correlacionan la información disponible sin constituir grupo de control ni manipular la información

El esquema es el siguiente:





GRÁFICA09: Esquema de diseño correlacional

Dónde:

n = Muestra

O_1 =: Detección de fraude en telefonía celular

O_2 = Utilización de las redes neuronales

r : Indica la relación entre las variables

5.9 Población y muestra

La población total considerada para esta investigación es la sumatoria de los elementos considerados para las llamadas LOC= 144; para las llamadas NAT = 64 y para las llamadas INT = 36; en total constituyendo una población y muestra de 244 elementos a las que se les aplicara las pruebas correspondientes, es decir, la muestra es igual a la población, lo que indica que se trabajara con el total de la población.

Tabla 3

Población considerada en la ciudad de Lima

Tabla No. 03

GRUP. CONSID.

TIPO DE LAM. LOC NAT INT. TOTAL

CANT./LLAM.	144	64	36	244
--------------------	------------	-----------	-----------	------------

La población y muestra tomada para este tipo de investigación representa a la población total tomada a priori, la misma que representa también a la muestra, porque se ha decidido trabajar con el total de dicha población pero si se debe tener en cuenta que los elementos tomados están directamente relacionados con el problema, es decir corresponden a las variables materia des del trabajo.

Técnicas de recolección de datos

Las técnicas aplicadas en este trabajo ha sido la elaboración de patrones de uso para cada tipo de llamada, es decir, LOC, NAT e INT, de tal forma que en el eje de las abscisas de la grafican se ha considerado la hora de las llamadas y en el eje de las ordenadas, la duración de cada llamada.

Se sabe que la metodología en todo proceso de investigación es una disciplina técnica que permite al investigador elaborar en forma sistemática el conjunto de métodos para que aplicando ciertas técnicas y procedimientos se logren los objetivos deseados. Por estas razones esta investigación posee un enfoque **cuantitativo**. Con respecto a este caso Hernández, Fernández y Baptista (2006), en su libro Metodología de la Investigación, sostienen: : que el método es conocido como Hipotético Deductivo, porque cumple las siguientes funciones:

- 1° Hace formulaciones del problema materia de la investigación.
- 2° Se plantean los objetivos en relación a los problemas previamente formulados.
- 3° Las variables que participan en la investigación se ubican en los problemas.
- 4° Los objetivos se analizan ampliamente dentro del marco teórico, pues de éste se extrae las hipótesis que antes se corresponden coherentemente a los problemas y objetivos previamente formulados.

a. Instrumentos empleados

Las herramientas utilizadas por el investigador es la captación a través de las redes neuronales las cantidades de llamadas antes mencionadas 'para cada tipo de llamada, luego administrar el uso de las llamadas y de esta manera detectar quién llama, en que modalidad, la duración, la hora y de esta forma detectar las posibles fugas en el uso de la línea.

En base a la definición dada a los instrumentos empleados, es importante mencionar que en los análisis de este trabajo se ha utilizado los patrones de usuarios en torno a las llamadas LOC, NAT e INT, relacionándolo con las variables Detección de fraude en telefonía celular y Utilización de las Redes Neuronas en la ciudad de Lima, en año 2015.

5-.10 Validez y confiabilidad

Validez

La autoría de los diferentes análisis realizados en esta tesis análisis corresponden al investigador: Bach. José Luis Aguilar Montesinos, trabajo que, para su validez de los ítems correspondiente a la primera variable, se ha recurrido a la opinión de expertos en el dominio de los temas que se relacionan con nuestra investigación. Para tal caso se ha recorrido al apoyo de Magísteres y Doctores, cuyos aportes han servido para mejorar el análisis y uso de los instrumentos, generando como resultado a opinión de los especialistas un resultado de 96.5% de aprobación.

Tabla 4

Validación de cuestionario de percepciones de La Detección de frauden en telefonía celular

N°	EXPERTO	CALIFICACIÓN	PORCENTAJE
1	Experto Augusto A. Matos R.	97	97.0%

Experto			
2	Arriola R. Franklin	95	95.0%
Experto			
3	Barraza T. Julio	96	96.0%
PROMEDIO TOTAL		288	96.00%

FUENTE: Opinión de expertos.

TABLA N° 04

En la tabla 4, se establece que el uso del instrumento dirigido a la población y/o muestra, de acuerdo con la opinión de los profesionales experimentado consultados, la aprobación se encuentra en 97.00%, el 95% para el segundo experto y el 96% para el tercero, generando un promedio del 96% de aprobación. Tasa que es el resultado de la utilización e instrumentos y los cálculos obtenidos, que del promedio de 81 a 100%, se ha logrado el 96%, valor de alta consideración.

5.6.2. Confiabilidad

En lo que respecta al proceso empleado para la determinación de la confiabilidad, se ha procedido a la aplicación de una prueba piloto, lo cual tiene como objetivo la obtención del Alfa de Cronbach, de los instrumentos empleados haciendo uso de la técnica de ítem-test, a través del mismo medio se ha podido medir la importancia de los cuestionarios: que constituyen las técnicas las técnicas de cambio de los patrones de uso en plena identificación de los sujetos en cada tipo de llamadas, con sus respectivos tiempos de duración de tales llamadas. De esta manera se obtuvo la confiabilidad partiendo de pruebas aplicadas a 244 usuarios que reunían deferentes característica y comportamientos en el uso de los servicios de telefonía dentro de los tres patrones considerados como para este tipo de análisis útiles de análisis. Por lo general en este tipo de análisis se recurre a ordenar las respuesta a los Items estructuradas por la escala de Liker, para el mismo siempre se le asigna valores.

Tabla 5

Escala de opiniones tipo Likert de la población de la Lima distrito a cerca de las percepciones de la Detección de fraude de telefonía celular

ESCALA	DESCRIPCION	SIGLA
1	Nunca	N
2	Casi nunca	CN
3	A veces	A V
4	Casi siempre	CS
5	Siempre	S

TABLA N° 05

Luego de procesar la información recolectada bajo este método, se procedió a determinar el coeficiente Alfa de Cronbach con respecto a la percepción de la Detección de fraude de telefonía celular, para lo que técnicamente se aplica sobre un baremo, con la finalidad de poderlo interpretar adecuadamente.

Tabla 6

Baremo del coeficiente Alfa de Cronbach				
-1 a 0	0,01 a 0,49	0,5 a 0,75	0,76 a 0,89	0,9 a 1
No es confiable	Baja Confiabilidad	Moderada confiabilidad	Fuerte Confiabilidad	Alta Confiabilidad

Finalmente, juego de las pruebas se calcula el coeficiente del Alfa de Cronbach.

Tabla: 07.

Cálculo del Coeficiente del Alfa de Cronbach del instrumento de medición de las variables de estudio

Instrumentos de medición	Alfa de Cronbach	N de elementos
Percepción de la Detección de fraude en telefonía celular y la utilización de las Redes Neuronales como medio de control	0.971	27

En esta prueba realizada complementariamente sobre 27 elementos del total de los considerados para el estudio, los valores encontrados para la cantidad de información en cada tipo de llamadas, arrojan para la prueba de Detección de fraude de telefonía celular un valor de 0.971, considerándose como de fuerte confiabilidad.

Tabla: 08

Interpretación de los valores de los coeficientes de correlación (teoría)

Coeficiente	Interpretación
0.0	Relación nula
0.00 – 0.200	Relación muy baja
0.200 – 0.400	Relación baja
0.400 – 0.600	Relación moderada
0.600 – 0.800	Relación alta
0.800 – 1.00	Relación muy alta
1.00	Relación perfecta

Finalmente, para efectuar todo este proceso se utilizó el programa estadístico SPSS versión

VI DISCUSIÓN DE LOS RESULTADOS

La investigación en este campo que presenta muchas dificultades en su proceso, realmente se puede considerar un trabajo bastante satisfactorio, desde el punto de vista en que hemos logrado determinar cambios de comportamiento en los usuarios investigados. A l margen que el cambio de comportamiento no necesariamente es considerado como actitud fraudulenta, pero determinar los métodos de análisis, para acotar a las investigaciones futuras sobre estos grupos de usuarios que tienen comportamientos diferentes a los normal, es decir, actitudes de fraude. Cuando se aplican otros métodos y técnicas, según (ASPeCT, 1997), también se pueden obtener un alto grado de veracidad, manifestado a través de usuarios que se encuentren haciendo uso de sus teléfonos celulares “fraudulentamente”.

Asimismo, las pruebas realizadas han servido para identificar a los clientes que cambian de comportamiento en el uso de telefonía celular de manera ilegal. Lo que se puede argumentar sobre ellos que se consideran usuarios que hacen uso de un alto nivel de consumo del servicio internacional, para luego realizar llamadas Locales. Desde el punto de vista comercial, representa una información bastante importante hacer las evaluaciones necesarias, debido a que por cualquier razón personal deciden cancelar el uso de teléfonos celular de llamadas INT, actitud que nos permiten concluir y crear nuevos y mejores planes de tarifas basadas en estos resultados. De la misma forma se cree que queda demostrado a través de las pruebas aplicadas y demostradas, donde el análisis diferencial siempre provee mucha información en comparación con el análisis absoluto; mediante el mismo es posible, y además nos permite reconocer a los usuarios en investigación.

Finalmente, a manera de concluir adicional, se puede sostener que, las redes neuronales nos demuestran que son una de las mejores herramientas clasificar las llamadas y construir los perfiles de usuario, debido a que representan a las herramientas más eficaces para medir y controlar el comportamiento de los usuarios en situaciones de duda de actitudes y comportamientos.

VII CONCLUSIONES Y RECOMENDACIONES

7.1 CONCLUSIONES:

1° caso.: Inicialmente se parte de esta cuestión ¿cuáles son las estructuras adecuadas que se les puede fijar para a los registros CUP y UPH?

Las estructuras orientadas a estos dos instrumentos y consideradas para nuestro análisis han sido los patrones de llamadas, para cada tipo de servicio, es decir LOS, NAT y los patrones de llamadas INT.; los mismos que han demostrado eficiencia en el control del consumo del usuario, en donde a través de los análisis y experimentos se puede determinar una idea con respecto al comportamiento del usuario, en lo referente al comportamiento actual e histórico. También sirve para tener conocimiento con respecto a la distribución de frecuencias, relacionándolo con el tipo de llamadas.

2° Caso: ¿En los casos que considerara una información correcta, cuántos prototipos deben tener los registros CUP y UPH?

A través de las pruebas respectivas se han hecho las definiciones de las estructuras con 144 patrones LOC, 64 NAT y 36 INT, determinando la suficiencia para analizar el comportamiento de los usuarios, tal como se resolvió en el primer caso.

3° Caso: Si tenemos ya los patrones definidos, ¿cuál debe ser la manera de clasificar las llamadas para el logro de los objetivos?

La utilización de las redes neuronales SOM nos han demostrado ser buenos ordenadores de las llamadas, tal como se puede observar en todos los análisis presentados, donde se puede notar que el espacio de llamadas se ha agrupado en forma discreta. Dichos grupos son bastante resaltantes en cuanto a su comportamiento de consumo de usuarios, teniendo en cuenta que los clientes de la empresa analizada, son en su mayoría son otras empresas y los clientes indirectos son los usuarios individuales, donde el mayor uso de los servicios se encuentra dentro de esta franja. Por lo cual, una vez que se ha definido la franja de 8.00 am – 20.00 horas, se indica que en este lapso se realiza la mayor parte de sus llamadas de empresas y no usuarios individuales, donde la frecuencia del mayor tiempo empleado en las llamadas se encuentra en la franja horaria laboral

de 8 a 20hs. con duraciones de entre 0 y 7 minutos). Una vez definidos estos patrones, cada llamada que ingresa al sistema, con cierta facilidad se le clasifica en el patrón que posee mayor claridad y también que con facilidad se adapte al perfil, presentado en la ecuación de Grábec, para de esta forma determinar el campo estocástico de las redes SOM. A demás es una forma de añadir información a los perfiles, lo que genera variación en la distribución de frecuencias, según sean los cambios de comportamiento observado.

4° Caso: La codificación de las llamadas, útiles para servir como prototipos

La información obtenida a través de las llamadas, observados a través de ciertos factores como el tipo de llamada, la hora y la duración. Como parte inicial se considera el tamaño o dimensión de la llamada y determinar, qué red neuronal se está utilizando para ordenarla adecuadamente y los dos siguientes pasos han definido el ingreso definitivo a la red neuronal correspondiente. En efecto se puede argumentar que, cuando la información presentada y codificada adecuadamente, sirven para construir los perfiles y definir los patrones correspondientes.

Caso 5: La interrogante que directamente tiene que ver con la investigación, es decir, ¿cómo detectar los cambios en el patrón de consumo de un usuario?

Se trata de que una vez que se ha definido el tamaño de las estructuras de los perfiles como vectores que reflejan que, cada valor representa una frecuencia del patrón a donde corresponde, también indica que cualquier distancia vectorial sirve para detectar diferencias entre los perfiles CUP y UPH. Sin embargo, basados en los planteamientos de los autores Burge y Taylor y la distancia de Hellinger (H), se puede determinar en forma correcta, cuándo realmente se presentan cambios de comportamiento.

7.2 RECOMENDACIONES:

1° Distinguir con bastante claridad dentro de los Registros CUP y UPH los patrones de los servicios LOC, NAT y INT, para que a través y de pruebas y experimentos observar el comportamiento de los usuarios, teniendo en cuenta que por medio de la distribución de frecuencias, determinar en cuál de ellos se ha dado la filtración.

2° Establecer que como cantidad mínima de elementos con que se puede trabajar en cada grupo son: LOC = 144, NAT = 64 y INT = 36 elementos, pudiendo ser mayor este número, pero los resultados serían los mismos; pero si la cantidad de elementos es menor, es bastante probable en que los resultados no sean satisfactorios, salvo utilizando métodos diferentes.

3° Se recomienda utilizar las Redes SOM, consideradas como excelentes ordenadores de llamadas, esta tiene la característica de clasificar y ordenar el tipo y hora y duración de la llamada, determinándolo y registrándolo en sus respectivos grupos.

Otro detalle las Redes SOM tienen la capacidad de agregar información a los perfiles, combinándolo según corresponda.

4° Clasificar las llamadas a través de la información obtenida, es decir, si son:

5. Determinar la llamada son: LOC, NAT o INT.
6. Establecer a qué hora se realiza dicha llamada
7. Considerar cuánto tiempo dura cada llamada para establecer promedios

Determinadas estas características observar la dimensión o tamaño del servicio a través de la Red Neuronal, valorándolo con una codificación que va desde (0 a 1), lo que nos permitirá reforzar los perfiles del usuario y definir los patrones correspondientes.

5° Determinar los perfiles como vectores para representar una frecuencia de cada patrón. Es importante establecer que cualquier distancia vectorial, en este caso, son útiles para hacer cálculos y establecer diferencias entre los perfiles CUP y UPH. Asimismo, se recomienda el uso de la distancia de Hellinger, debido a que este método es bastante exacto en el estudio de los cambios de comportamiento de los usuarios.

REFERENCIAS:

- 1.- ASPeCT, 1999. Definition of Fraud Detection Concepts, Deliverable D06 Pág.47
- 2.- ASPeCT, 2000. Fraud Management tools: First Prototype, Deliverable D08.Pág.31
- 3.- ASPeCT, 1997. Fraud Detection Concepts: Final Report, Deliverable D18, Pag.27
- 4.- Burge P., Shawe-Taylor J., 1996. Frameworks for Fraud Detection in Mobile Telecommunications Networks, Department of Computer Science Royal Holloway, University of London.
- 5.- Burge P., Shawe-Taylor J., 1997. Detecting Cellular Fraud Using Adaptive Prototypes, Department of Computer Science Royal Holloway, University of London

6.- Hilera J. R., Martínez V. 2000. Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones, RA-MA Editorial, Madrid.

7.- Banzhaf W., 2000, Self-organizing Systems. Dept. of Computer Science University of Dortmund Dortmund Germany Informatik Centrum Dortmund

8.- Hollmen J., Tresp V. Call-based Fraud Detection in Mobile Communication Networks using a Hierarchical Regime-Switching Model, Helsinki University of Technology Department of Computer Science and Engineering Laboratory of Computer and Information Science. Siemens A. G., Corporate Technology.

9.- Hollmen J, 2004. Process modeling using the Self-Organizing Map, Helsinki University of Technology Department of Computer Science

10.- Burge P, Shawe-Taylor J. Fraud Detection and Management in Mobile Telecommunications networks, Department of Computer Science Royal Holloway, University of London. Vodafone, England. Siemens A. G

11., Burge P., Shawe-Taylor J., 2000. Detecting Cellular Fraud Using Adaptive Prototypes, Department of Computer Science Royal Holloway, University of London.

12,- Burge P., Shawe-Taylor J., 2002. Fraud Detection and Management in Mobile Telecommunications networks, Department of Computer Science Royal Holloway, University of London. Vodafone, England. Siemens A. G. Proceedings of the 2nd European Conference on Security and Detection, IEE Conference publication 437, pp. 91-96, London

ANEXOS:

PRESUPUESTO DE LA INVESTIGACIÓN

Costos de la investigación

NECESIDAD	CANTIDAD	COSTO UNITARIO	TOTAL COSTO
Recursos Humanos			

1.- Asesores	2	1,500.00	3,000.00
2.- Auxiliar de Investigación	2	1,000.00	2,000.00
3.- Operadores de la información	3	600.00	1,800.00
4. Tipadoras	1	800.00	800.00
Recursos Materiales			
1.- Movilidad y viáticos	4	400.00	1,600.00
2.- Servicio de computo e Impresión	Varios	900.00	900.00
3.- Adquisición de una impresora	01	800.00	800.00
4.- Adquisición de libros, Revistas y otros medios	Varios	1,300.00	1,300.00
USV y otros medios de Almacenamiento de Datos	2	35.00	70.00
Papel en general	4millares	13.50	54.00
Tinta de Impresora, lápices y lapiceros	Varios	150.00	150.00
Scaneos y otros materiales especiales	Varios	100.00	100.00
Fotocopia	4 millares	0.05	200.00
Anillados	15 Unidades	7.00	105.00
TOTAL			12,879.00

ASPECTOS ADMINISTRATIVOS

Cronograma y costos del estudio

CÓDIGO	ACTIVIDADES	2015
---------------	--------------------	------

Matriz de Consistència

TÍTULO: Detección del problema de fraude en telefonía celular, con el uso de redes neuronales en la ciudad de lima, año 2015.

Autor(a): Br. José Luis Aguilar Montesinos

TÍTULO:	Objetivos	Hipótesis	Variables e indicadores						
<p>Problema General:</p> <p>¿De qué manera se relacionan la Detección de fraude en el uso de telefonía celular con la Utilización de las Redes Neuronales en la ciudad de Lima, periodo 2015?</p> <p>Problemas Específicos:</p> <p>De qué manera se relacionan la Detección del fraude en el uso de redes neuronales con</p>	<p>Objetivo general:</p> <p>Determinar en forma exacta la relación existente entre la detección de fraude en el uso telefonía celular con la utilización de las Redes Neuronales en la ciudad de Lima, año 2015.</p> <p>Objetivos específicos:</p> <p>Establecer la relación que mantienen el fraude en el uso de las redes neuronales con los costos de los servicios de telefonía en la ciudad de Lima, periodo 2015.</p> <p>Determinar la relación entre la detección de fraude en el uso de</p>	<p>Hipótesis general:</p> <p>H_a. Se ha determinado que existe una relación significativa entre la detección de fraude en el uso de telefonía celular con la utilización de las Redes Neuronales en la ciudad de Lima, periodo 2015.</p> <p>Hipótesis específicas:</p> <p>H₁. Se ha podido establece un alto grado de relación entre el fraude en el uso de telefonía celular con los costos de los servicios de</p>	Variable I: El Código Penal Peruano						
			Dimensiones	Indicadores	Ítems			Escala de medición	Niveles o rangos
			<p>Costos de los servicios de telefonía</p> <p>Estructuras de patrones de llamada LOC, NAT y INT</p>	<p>-Las técnicas empleadas en el robo de línea</p> <p>-Los daños cometidos en los usuarios legales.</p> <p>- La frecuencia con que se comete el robo.</p> <p>-Los reclamos sin solución al problema</p> <p>- Los montos elevados que están fuera del de los presupuestos</p> <p>-El abandono al uso del servicio.</p>	<p>1 2 3</p> <p>4 5 6</p> <p>7 8 9</p> <p>10 11 12</p> <p>13 14 15</p> <p>16 17 18</p>	<p>Nunca</p> <p>Casi Nunca</p> <p>A Veces</p> <p>Casi Siempre</p>	<p>No Preocupante</p> <p>(8101-135)</p> <p>Medio</p> <p>(64-100)</p> <p>Preocupante</p> <p>(27- 63)</p>		

<p>los costos de los servicios de telefonía celular en la ciudad de Lima, año 1015?</p> <p>¿Qué grado de relación existe entre el uso de las redes neuronales con la estructura de patrones LOC, NAT y INT en la ciudad de Lima, periodo 2015?</p> <p>¿Cuál es la relación que se observa entre el uso de las redes neuronales con la construcción y mantenimiento de los perfiles del usuario en la ciudad de Lima, periodo 2015?</p>	<p>las redes neuronales y las estructuras de patrones de llamadas LOC, NAT y INT en la ciudad de Lima, año 2015.</p> <p>Establecer la relación el fraude en el uso de las redes neuronales y la construcción y mantenimiento de los perfiles del usuario en la ciudad de Lima, año 2015.</p>	<p>telefonía en la ciudad de Lima, año 2015.</p> <p>H₂. Existe una relación bastante considerable entre la detección de fraude en el uso de la telefonía celular y las estructuras de patrones de llamada LOC, NAY y INT en la ciudad de Lima, año 2015.</p> <p>H₃. Los estudios demuestran que existe una relación determinante entre el fraude en el uso de las redes neuronales y la construcción y mantenimiento de los perfiles del usuario en la ciudad de Lima, periodo 2015.</p>	<p>Construcción y mantenimiento de los perfiles del usuario</p>	<p>-Medición de la rentabilidad entre los tipos de llamadas.</p> <p>Los costos de mantener bien los perfiles.</p> <p>-Los beneficios obtenidos.</p>	<p>19 20 21</p> <p>22 23 24</p> <p>25 26 27</p>	<p>Siempre</p>	
Variable D:El delito de la pornografía infantil							
Dimensiones		Indicadores		Ítems		Escala de valores	Niveles o rangos
El robo de línea como elemento característico del fraude		<p>-El comportamiento de la población frente al creciente delito</p> <p>-Acción de los organismos privados frente a estos problemas</p> <p>-El comportamiento de los alumnos adolescentes frente a la publicación de estos delitos</p> <p>Los reclamos sin solución al problema.</p>		<p>28 29 30</p> <p>31 32 33</p> <p>34 35 36</p> <p>37 38 39</p>		<p>Nunca</p> <p>Casi Nunca</p> <p>A Veces</p>	<p>No Preocupante</p> <p>(8101-135)</p> <p>Medio</p> <p>(64-100)</p> <p>Preocupante</p> <p>(27- 63)</p>

			<p>El comportamiento de los afectados por el robo de línea</p> <p>El efecto social que genera este problema.</p>	<p>-Los montos elevados que están fuera del de los presupuestos</p> <p>El abandono al uso del servicio.</p> <p>- Pérdida de ingresos del fisco por impago de impuestos.</p> <p>--El incremento del fraude por esta modalidad.</p> <p>- El descontrol por parte de las instituciones</p>	<p>40 41 42</p> <p>43 44 45</p> <p>46 47 48</p> <p>49 50 51</p> <p>52 53 54</p>	<p>Casi Siempre</p> <p>Siempre</p>	
Tipo y diseño de investigación	Población y muestra	Técnicas e instrumentos		Estadística a utilizar			
<p>Tipo: Descriptiva</p> <p>Alcance</p> <p>Diseño: Correlacional</p>	<p>Población: 244</p> <p>Personas</p>	<p>Variable I: Detección del fraude en telefonía</p> <p>Técnicas: Pruebas técnicas</p> <p>Instrumentos: Patrón de usuarios</p> <p>Autor: Br. José Luis Aguilar Montesinos</p> <p>Año: 2015</p> <p>Monitoreo:</p>		<p>DESCRIPTIVA:</p>			

Método: Deductivo	Hipotético	Tipo de muestreo: Probabilístico o	Ámbito de Aplicación: La ciudad de Lima Forma de Administración: Personal	INFERENCIAL: X
		Aleatorio	Variable D: Fraude en telefonía Celular	
		Tamaño de muestra: 244 personas	Técnicas: Pruebas técnicas	
			Instrumentos: Técnicas de control	
			Autor: Br. José Luis, Aguilar Montesinos Año: 2015 Monitoreo: Ámbito de Aplicación: La ciudad de Lima Forma de Administración: Personal	

