REVISTA CUBANA
DE TRANSFORMACIÓN DIGITAL

RNPS 2487 • ISSN 2708-3411
Vol. 3 • Nro. 4 • octubre-diciembre 2022 • e182

ARTÍCULO ORIGINAL

Métodos y Técnicas de *Machine Learning* e Inteligencia Artificial para el enfrentamiento al Fraude en las Telecomunciaciones

Machine Learning and Artificial Intelligence Methods and Techniques to combat Fraud in Telecommunications

Claudia Beatriz Martínez Castro
claudia.martinez@etecsa.cu • https://orcid.org/0000-0002-0116-8396

EMPRESA DE TELECOMUNICACIONES DE CUBA, ETECSA, CUBA

Jose Alberto Vilalta Alonso
jvilalta@ind.cujae.edu.cu • https://orcid.org/0000-0001-7505-8918

UNIVERSIDAD TECNOLÓGICA DE LA HABANA "JOSÉ ANTONIO ECHEVERRÍA", CUJAE, CUBA

Recibido: 2022-09-23 • Aceptado: 2023-12-25

RESUMEN

Este trabajo recoge un estudio bibliográfico sobre diferentes métodos y técnicas de minería de datos (MD), *Machine Learning* (ML) o aprendizaje automático, e Inteligencia Artificial (IA), asociados al enfrentamiento al fraude en las telecomunicaciones, el cual está en constante transformación y se ha complejizado a la par que los servicios y las tecnologías. Cada vez la cantidad de datos para procesar es mayor, lo que incide en un aumento del tiempo de respuesta al fraude si no se emplean técnicas apropiadas, además de que se exige la combinación de diversas fuentes de datos, por lo que este tipo de herramienta es fundamental, tanto para la detección de patrones (comportamientos de fraude), como para la automatización de los procesos de trabajo que permitan reducir los tiempos de respuesta. Esto se logra con la aplicación de una serie de métodos que pueden ser supervisados, semisupervisados y no supervisados, que comprenden algoritmos para el tratamiento de grandes volúmenes de datos, específicamente para el tratamiento del fraude *bypass*. Al reducir el tiempo de detección y mitigación del fraude, así como la correcta caracterización de patrones de comportamiento fraudulentos, se garantiza el aseguramiento de los ingresos y se evitan pérdidas económicas.

Palabras clave: aprendizaje automático, fraude en telecomunicaciones, fraude *bypass*, Inteligencia Artificial, minería de datos.

ABSTRACT

The present work includes a bibliographical study on different methods and techniques of Data Mining (MD), Machine Learning (ML) and Artificial Intelligence (AI), associated with the fight against fraud in telecommunications, which is in constant transformation and has been as the services and technologies become more complex, the amount of data to be processed is increasing, which leads to an increase in the response time to fraud if appropriate techniques are not used, in addition to requiring the combination of various data sources, so this type of tool is essential, both for the detection of patterns (fraud behaviors) and for the automation of work processes that allow reducing response times to it, this is achieved with the application of a series of methods that can be supervised, semi-supervised and unsupervised that comprise a series of algorithms for the treatment of large volumes of data, and specifically for bypass fraud treatment. By reducing fraud detection and mitigation time, as well as the correct characterization of fraudulent behavior patterns, income assurance is guaranteed and economic losses are avoided.

Keywords: *Machine Learning, telecommunication fraud, bypass fraud, artificial intelligence, data mining.*

INTRODUCCIÓN

El enfrentamiento al fraude en las telecomunicaciones es un fenómeno complejo de constante lucha entre los analistas de fraude y los defraudadores; depende de las características de los servicios y las tecnologías, así como de la evolución de estos, medios de pago utilizados, terminales, entre otros. Además, este fraude provoca daños en las empresas además de financieros de imagen y credibilidad. Es fundamental los medios y las tecnologías empleados para su detección, neutralización y minimización de su impacto (Pérez *et al.* 2004).

El fraude siempre ha tenido un impacto significativo en las pérdidas de las empresas, según datos publicados en 2021 por la Asociación de Control del Fraude en las Comunicaciones (CFCA por sus siglas en inglés). Se conoce que en 2019 las pérdidas globales por concepto de fraude ascendieron a 28,3 miles de millones de dólares en el último período medido

(2017-2019), de los cuales 2,71 miles de millones de dólares se relacionan con el fraude *bypass*.

La gestión del fraude está estrechamente relacionada con el aseguramiento de ingresos. Esta tiene sus inicios, como área de conocimiento, precisamente en el sector de las telecomunicaciones, a finales de la década de 1970. La actividad de aseguramiento de ingresos está enfocada a definir y eliminar las causas técnicas y estructurales relacionadas con la fuga de ingresos (Castro, 2017), por lo que es fundamental conocer las brechas de seguridad en los diferentes servicios de telecomunicaciones, así como de las deficiencias técnicas; para ello se hace necesario realizar un monitoreo exhaustivo de estos, atendiendo por supuesto a sus particularidades y según el tipo de red.

En la actualidad, 45 % de las empresas no emplean procesamiento automático de datos, contra un 23 % que sí lo hace, mientras que 15 % tenía planificado la introducción de este tipo de procesamiento y 18 % no estaba seguro de su aplicación. Entre las diferentes herramientas para el enfrentamiento al fraude, por la CFCA, se encuentra el procesamiento manual, que es aplicado por: 30% de las empresas encuestadas y 28 % de Sistemas de Gestión del Fraude (FMS). Cuba utiliza ambos: 15 % motores de decisión y 13 % emplea técnicas y métodos de *Machine Learning* (ML) o aprendizaje automático, e Inteligencia Artificial (AI), por lo que se apuesta en la actualidad y lo que se necesita potenciar debido a los buenos resultados obtenidos en estudios realizados en otras partes del mundo; queda 12 % que contrata el servicio a expertos y 2 % que hace uso de otros métodos (CFCA Report, 2021).

Por lo tanto, en este trabajo se pretende hacer una revisión bibliográfica de materiales basados en estudios que se realizaron en otras empresas de telecomunicaciones y en universidades donde se aplican métodos y técnicas de ML e IA, de manera general, y que pueden ser aplicados en el enfrentamiento al fraude *bypass*.

METODOLOGÍA

Se realizó un estudio del arte sobre diferentes aplicaciones de la minería de datos (MD), *Machine Learning* (ML) y la Inteligencia Artificial (IA), en el enfrentamiento al fraude en las telecomunicaciones, para llegar concretamente a la aplicación del fraude *bypass*. Para esta búsqueda, se utilizó principalmente el Google académico, empleando términos de búsqueda y operadores lógicos basados en palabras claves de interés.

El gestor bibliográfico empleado fue el EndNote X7 y se prestó atención a las publicaciones realizadas en inglés, idioma más generalizado en este tipo de literatura, esencialmente de 2017 a 2022. Coinciden en este porcentaje los artículos de revistas, el resto son tesis, actas de congresos, secciones de libros, etcétera.

Se extrajo la información sobre las aplicaciones realizadas en cuanto a diferentes tipos de procesamiento, así como de la eficiencia y eficacia de los algoritmos aplicados en los casos expuestos, los cuales se muestran a lo largo de este artículo.

La introducción de estas técnicas y herramientas produce siempre un impacto en las organizaciones, por lo que es necesario, en primer lugar, modelar los procesos de gestión de la

actividad y definir los factores que inciden en la gestión del fraude *bypass*; indicadores; métodos; técnicas y tecnologías empleadas; así como la caracterización de los elementos referentes a cada factor. Se deben definir las fuentes de datos que alimentarán los modelos, caracterizar las variables que corresponden a indicios y describir tanto los elementos de tráfico como comerciales asociados al servicio móvil, que se conocen y son importantes para su gestión.

RESULTADOS Y DISCUSIÓN

Existen diferentes tecnologías que se usan para el estudio del fraude en esta industria, como son las técnicas estadísticas para la extracción de características, realizar clasificaciones y llevar a cabo un procesado temporal de los datos, todo basado en reglas donde se modelan y configuran patrones de fraude basados en umbrales de detección y, por último, las técnicas de aprendizaje automático que parten de un entrenamiento para realizar la detección. En el caso de estas últimas se pueden emplear diferentes modelos, como redes neuronales supervisadas y realimentadas, muy usadas en los fraudes asociados a móviles. Está generalizado el uso de esquemas híbridos.

Como se ha mencionado con anterioridad, en el enfrentamiento al fraude en las telecomunicaciones es esencial la rapidez con la que se actúe sobre el evento fraudulento. Específicamente en el caso del fraude *bypass* tiene una alta connotación. Si además esto se une a la necesidad de analizar un conjunto considerable de datos de servicios que tienen un comportamiento sospechoso, con todas las variables asociadas a esos servicios, y qué incidencia tienen esos datos en la explicación del fenómeno, ante estas dificultades se precisa la aplicación de la minería de datos (MD), con el fin de extraer al máximo la información útil dentro de esa gran cantidad de datos, que además provienen de diferentes fuentes de información.

Se hace necesario definir lo que es MD. Mac Lennan en 2008 planteó que es el proceso de analizar datos usando metodologías automatizadas para encontrar patrones escondidos, que se aplican a los datos generados por un proceso o negocio para obtener información que soporte la toma de decisiones, lo que se alcanza con la automatización del proceso de encontrar información predecible en grandes bases de datos y resulta la respuesta a preguntas que se procesaban de manera manual (Menes, 2015). Unos años después, en 2016, Frand plantea que es el proceso de análisis de datos desde diferentes perspectivas, resumiéndola en información útil para descubrir conocimiento (Albougha, 2016). Es un proceso interactivo e iterativo, resumido en cinco etapas: Selección, Pre-Procesamiento/limpieza, Transformación/reducción, minería de datos e Interpretación/evaluación.

El *Machine Learning* o aprendizaje automático está fuertemente relacionado con la minería de datos; pero incluye algoritmos para el aprendizaje, a partir de datos históricos y de lograr la automatización de determinados análisis. Es un subconjunto de la Inteligencia Artificial, una parte de lo que se conoce como ciencias de la computación, término empleado para el análisis de información, el cual constituye una fracción importante de la analítica de datos. Se plantea que es pieza clave de la revolución del *Big Data*, que trabaja con un enfoque

estadístico y ayuda a predecir modelos de datos, al tiempo que no necesita pruebas estadísticas exactas (Arun, Ali, Irtaza y Anwar, 2020).

La medición del desempeño de estos algoritmos, según el tiempo de respuesta, el uso de recursos de cómputo y la confiabilidad de las operaciones, o sea, la precisión con la que un modelo define el conjunto de datos de entrada, es fundamental para cualquiera de las técnicas que se pueden aplicar en este campo (Menes, 2015).

Se identifican dos vertientes fundamentales en el enfrentamiento al fraude *bypass*, la detección de patrones y la automatización del proceso de clasificación de los servicios sospechosos, mediante el aprendizaje automático. Los métodos de aprendizaje se pueden clasificar en tres tipos: supervisados, semisupervisados y no supervisados (Da Costa, 2020).

En el aprendizaje supervisado, que es el enfoque más empleado, se requiere partir de un conjunto de datos donde exista una definición previa de etiquetas, por ejemplo, «fraude» o «no fraude, lo cual es imprescindible para el entrenamiento de un clasificador. Su principal ventaja es que las salidas de este tipo de algoritmo son significativas para el entendimiento humano. También tiene limitaciones, partiendo de lo difícil que es recolectar las etiquetas, debido al volumen de datos y que se depende en buena medida del análisis correcto y experticia de los analistas de fraude; también siempre existen inconsistencias y ambigüedades que resultan complicadas a la hora de definir las etiquetas, lo que puede obstaculizar su implementación (Albougha, 2016). Dentro de los algoritmos de clasificación existen redes neuronales, vecino más cercano K, árboles de decisión, regresión logística, Naive Bayes y la técnica de máquina de soporte de vectores (SVM), Random Forest, entre otros. Otro tipo de algoritmos que entran en este grupo son los de regresión lineal, simple y logística. Todo depende del tipo de variable de salida.

Ighneiwa y Mohamed, en 2017, publicaron un trabajo sobre la aplicación de un enfoque basado en IA, a raíz de un estudio aplicado en Libia con el operador de móviles Tier 1, para enfrentar el fraude *bypass*, donde se referencia una serie de trabajos sobre el enfrentamiento a este tipo de fraude, combinando el análisis de los Detalles de Registros de Llamadas (CDR, Call Detail Records), con algoritmos de ML, donde se emplean varios clasificadores, como se puede observar en la figura 1, mostrando una gran efectividad al verificar la exactitud de estos (Ighneiwa, 2017).

Tabla 1: Varios estudios que emplean algoritmos clasificadores en la detección del fraude (elaboración del autor).

Estudios	Clasificadores	Exactitud
R.S.A.H. Elmi, S. Ibrahim "Detecting SIM Box Fraud Using Neural Network"	Redes Neuronales Artificiales (ANN)	98,70 %
R. Salehuddin, S. Ibrahim, A. Mohd Zain, and A. Hussein Elmi, "Classification of SIM Box Fraud"	Máquina Soportada en Vectores (SVM)	99,60 %
	Redes Neuronales Artificiales (ANN)	98,69 %
I Murynets, M. Zabarankin, R. P. Jover, and A. Panagia, "Analysis and detection of SIMbox fraud in mobility networks"	Combinación linear de 3 clasificadores: Arboles de Decisión Alternados, Arboles Funcionales y Random Forest (Bosques Aleatorios)	99,95 %

Las técnicas de aprendizaje no supervisado detectan fraude en un conjunto de datos no etiquetados, o sea, no existe ningún tipo de etiqueta para la construcción del modelo, y su principal ventaja es que no se basan en una identificación precisa de los datos de la etiqueta, que a menudo tiene fundamentos escasos e inexistentes. Dentro de este grupo se encuentran algoritmos de agrupamiento, que es un método en el cual los datos se unen en grupos o clústers, según la similitud de sus características (Albougha, 2016). Muchas veces la definición está asociada al cálculo de distancia entre los vectores como K-Media, que es de los más conocidos. También hay algoritmos de reducción de dimensionalidad, como Análisis de Componentes Principales (PCA), asociación donde el objetivo es descubrir la probabilidad de la co-ocurrencia de elementos en una colección, por ejemplo, las reglas de asociación utilizadas para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

Por otro lado, las técnicas de aprendizaje semisupervisado se presentan cuando se tiene un conjunto de datos con una pequeña cantidad de etiquetas y un gran número de muestras no etiquetadas, y su principal objetivo es entrenar un clasificador para ambos tipos de datos etiquetados y no etiquetados, por lo tanto, tendrán mejor desempeño (Abdallah, Maarof y Zainal, 2016).

En los últimos tiempos se han realizado estudios del aprendizaje conjunto y está probado que la combinación de modelos mejora la actitud de las predicciones, los modelos híbridos han aumentado su aplicación (Getahun, 2020).

Es importante mencionar conceptos de dos capacidades de ML: aprendizaje distribuido y AutoML. El primero es una solución a gran escala, cuando existen limitaciones de memoria y procesamiento y se usan múltiples computadoras o procesadores multinúcleo en paralelo, cada uno de los cuales procesa un algoritmo de ML diferente o una porción de los datos; el segundo se centra en la falta de personal especializado, por lo que es necesario crear facilidades para que personas con conocimientos limitados en la materia puedan seleccionar el mejor modelo de ML posible, y es donde entra el AutoML (Ferreira, Pilastri, Martins, Santos y Cortez, 2020). Estas soluciones son fundamentales para la utilización óptima de los recursos, mejorar el desempeño de los algoritmos que se apliquen y que se generalice su utilización de manera que no sea necesario emplear personal especializado, del cual no se dispone para realizar estos análisis.

También existen las técnicas de detección de anomalías, que como indica su nombre y ya se había mencionado, permite identificar patrones de comportamiento en los datos que difieren del resto o de su comportamiento usual. Es necesario aclarar que un comportamiento anómalo es una forma de actuar no usual y si estas se emplean para violentar lo establecido, se reconoce como fraude, ataque, amenaza y violación de seguridad, entre muchos otros términos aplicables. Existen tres técnicas básicas para detectar y clasificar comportamientos anómalos (Eng, Garay y Martínez, 2018):

1. Técnicas basadas en estadísticas: definen un comportamiento estocástico (comportamiento sometido al azar y que será objeto de un análisis estadístico).

2. Técnicas basadas en el conocimiento: disponibilidad de conocimientos previos a los datos, que brindan robustez, flexibilidad y escalabilidad.
3. Técnicas basadas en aprendizaje automático (*Machine Learning*): permite una categorización de patrones, lo que brinda flexibilidad en la captura de interdependencias.

También se puede hablar de la detección del fraude basada en el mal uso, donde se establece lo que es un comportamiento fraudulento y fuera de ahí otros comportamientos definidos como normales. Para ello se emplean métodos basados en reglas, estadísticas o enfoques heurísticos, y así determinar la ocurrencia de un evento sospechoso. La detección del fraude es un sistema de expertos, considerado un mecanismo de detección simple y rápido, que se encuentra limitado porque no se pueden detectar nuevos patrones de fraude, ya que solo actúa sobre patrones conocidos.

Se tiene además el enfoque híbrido de detección de mal uso y anomalías, propuesto por algunos investigadores para obtener resultados óptimos, ya que esa detección cubre la deficiencia del mal uso de no detectar nuevos fraudes, mientras que la detección de anomalías no tiene la capacidad de generalización en reglas o reportes que definan comportamientos de fraude y la presencia de muchas falsas alarmas (Abdallah, Maarof y Zainal, 2016). Esta idea la respaldan Getahun (2020) y Kaiafas (2020).

Más recientemente se habla de las aplicaciones de transmisión que imponen restricciones únicas y obligan al análisis de una continua secuencia de datos que se obtienen en tiempo real o casi real, y en ocasiones limita el análisis de información por lotes e introduce una serie de variables que se deben considerar, en cuanto al diseño de algoritmos de aprendizaje automático, como los recursos de cómputo en términos de memoria, poder de procesamiento de la Unidad de Procesamiento Real (CPU) de la computadora y la banda ancha de comunicación. Por lo tanto, propone un desafío para la detección de anomalías en tiempo real, por lo que se propone una serie de algoritmos, como los de agrupamiento de flujos (Stream KM++, CluStream, ClusTree, DenStream y CobWeb) (Da Costa, 2020).

Como hemos visto, es cada vez más importante lo que se conoce como análisis de datos o analítica de datos, proceso de análisis multidisciplinario, sistemático y computacional, que se encarga de extraer valor y conocimiento de los datos. En las telecomunicaciones ha existido un crecimiento explosivo del tráfico, cuyas causas radican principalmente en el aumento de los usuarios en los servicios de esta industria. Las fuentes de datos en el sector cada vez son más numerosas y variadas; ya no solo se usan por personas, sino también por dispositivos que se comunican e intercambian datos entre sí. Estas tecnologías se identifican como Internet de las Cosas (IoT), dando pie a la llamada Cuarta Revolución Industrial y el tercer factor que incide en este crecimiento lo constituyen los servicios de comunicación que evolucionan y desarrollan constantemente, y que generan nuevas tecnologías empleadas por los usuarios como la nube, la realidad virtual, la Inteligencia Artificial y el video de «ultra alta resolución», por lo tanto, se posibilita la introducción del concepto *Big Data*, lo cual es un ecosistema complejo y amplio que abarca una gran variedad de áreas técnicas y que debe solucionar las limitaciones

de las tecnologías tradicionales, para procesar datos que cumplen con las cinco V: Volumen, Variedad, Velocidad, Veracidad y Valor, que contienen las herramientas indicadas para realizar este procesamiento (Orbe, 2018).

El reto de las técnicas actuales para la prevención y el enfrentamiento al fraude en las redes móviles, sigue siendo las fuentes de datos, ya que la información de los suscriptores proviene de diferentes flujos, por lo que se habla en términos de tener un único y centralizado sistema de gestión del fraude, lo que da pie al concepto de Sistemas de Gestión del Fraude o Sistemas Antifraude (Tarmazakov, 2018). Hoy en día muchos de estos sistemas, a pesar de recibir diferentes flujos de datos provenientes de diversas fuentes, operan sobre cada una de ellas de manera individual, por lo que es necesario pensar, al observar el desarrollo de los servicios de telecomunicaciones, en la combinación de dichas fuentes para configurar reglas y crear modelos analíticos.

En cuanto a los sistemas empleados en la gestión del fraude, se conoce que en la década del 2000 se comenzaron a desarrollar e implementar los Sistemas de Detección de Fraude (SDF), fundamentados en la construcción de reglas que se basaban en la experticia de los analistas. Posteriormente, se comenzaron a insertar en algunos sistemas los métodos de minería de datos, ante la complejización del fenómeno del fraude para lograr un enfrentamiento más efectivo. Entre las ventajas que presentan estos sistemas están:

1. Los patrones de fraude se obtienen automáticamente de los datos.
2. Especifican la probabilidad de fraude en cada caso, lo cual permite establecer prioridades.
3. Revela nuevos tipos de fraude que no se hayan definido con anterioridad.

Es común que los sistemas antes mencionados integren detección basada en anomalías y que se asienten sobre métodos para perfilar comportamientos; de esta manera se realiza un monitoreo en busca de cualquier desviación de lo que se considera normal, para lo cual se emplean métodos de aprendizaje supervisado, semisupervisado y no supervisado.

Un Sistema de Gestión del Fraude (FMS, *Fraud Management System*) es una herramienta que permite desarrollar las actividades que intervienen en la gestión del fraude, de manera interrelacionada y efectiva, en dependencia del uso de sus capacidades. Esta herramienta debe permitir el análisis y la visualización de datos de los diferentes servicios de telecomunicaciones, modelado de diferentes tipos de reglas —ya sean simples, estadísticas o de patrones inteligentes—, así como la incorporación de módulos de *Machine Learning* e Inteligencia Artificial, que se ajustarán según el tipo de servicio. Un FMS analiza los CDR para crear perfiles de uso conocidos como patrones o comportamientos de fraude que distingan a los usuarios normales de los fraudulentos y permite que se modelen en reglas, a partir de las cuales se obtendrán alarmas de servicios sospechosos de cometer fraude (Ighneiwa, 2017),

Algo que se debe tener en cuenta para enfrentar el fraude, es un sistema de información robusto, que permita gestionar el fenómeno de manera integral y efectiva. (Arana Porlles, 2017).

CONCLUSIONES

A partir de los años 2000, parte de una comunidad de investigadores relacionados con la ciencia de datos, viene fortaleciendo la aplicación de la analítica de datos para la solución de algunos problemas, como el análisis de riesgo y la detección del fraude de diversos tipos, mostrando una amplia divulgación de soluciones de aprendizaje automático para la detección de patrones de comportamiento, independientemente de herramientas para su enfrentamiento, desde el punto de vista técnico y de estructura de los servicios. Para el caso específico de fraude *bypass* es fundamental la reducción del tiempo de respuesta, con el objetivo de lograr una mitigación temprana de las pérdidas, por lo que se reconoce que la MD y las técnicas de ML e IA son fundamentales para la extracción de conocimiento, la detección de patrones, abarcando diferentes fuentes de datos, y la automatización de procesos de trabajo, con el uso de clasificadores que estarían entre las técnicas de aprendizaje supervisado.

La mayor parte de las fuentes consultadas aseguran que lejos de utilizar un algoritmo único para la modelación de patrones fraudulentos, lo más efectivo resulta emplear enfoques híbridos, algoritmos combinados que se ajusten a las diferentes fuentes ~~—o~~ conjunto de ellas~~—~~, para conformar un modelo óptimo, y proponen una serie de variables para medir el desempeño de estos modelos, siendo las más utilizadas: la exactitud del modelo, la cantidad de falsos positivos, el tiempo de respuesta y los recursos de cómputo empleados. También se resalta la importancia de los sistemas que se empleen en la gestión del fraude, en general, y particularmente del *bypass* y la integración de estos, que incluyan la posibilidad de extraer y explorar datos sobre los cuales aplicar diferentes técnicas de estadísticas y MD, así como la inclusión de módulos de ML e IA.

REFERENCIAS

Artículos de revistas

Abdallah, A., M. A. Maarof & A. Zainal (2016). "Fraud detection system: A survey". *Journal of Network and Computer Applications,* 68: 90-113.

Al Bougha, M. R. (2016). "Comparing data mining classification algorithms in detection of simbox fraud".

Alcívar León, C. R. (2020). "Diseño de un modelo predictivo a través de la técnica de minera de datos 'Random Forest'para la detección de fraude bypass en redes telefónicas en el Ecuador".

Arana Porlles, D. A. (2017). "Influencia de un sistema integrado de información en la gestión del fraude en una empresa de telecomunicaciones, 2016".

Arias, F. & N. Cerpa (2008). "Extendiendo el modelo e-SCARF de detección de fraude en sistemas de comercio electrónico". *Ingeniare. Revista chilena de ingeniería,* 16(2): 282-294.

Arun, K., M. Ali, G. Irtaza & T. Anwar (2020). "Big Data analytics with Machine Learning for iot sensed data". *Pal Arch's, Journal of Archaeology of Egypt/Egyptology* 17(7): 5293-5301.

Association, C. F. C. (2021). Fraud loss survey report, 2021.

Cano Cancela, J. R. I., David; Moreno Izquierdo, Raúl (2011). Improving Fraud Detection Modeling.

Castro Aguilar, G. F. (2017). "Modelo para el aseguramiento de ingresos en organizaciones orientadas a proyectos basado en minería de datos anómalos".

Chouiekh, A. & E. H. I. E. Haj (2018). "Convnets for fraud detection analysis". *Procedia Computer Science,* 127: 133-138.

Da Costa, R. F. L. (2020). "Next Generation Machine Learning Based Real Time Fraud Detection".

Eng, D. A. & S. G. G. Álvarez (2018). "Lógica Difusa para la detección de comportamientos anómalos en los sistemas de telefonía". *Tono, Revista Técnica de la Empresa de Telecomunicaciones de Cuba S.A.,* 14(2): 73-84.

Ferreira, L., A. Pilastri, C. Martins, P. Santos & P. Cortez (2020). "An automated and distributed Machine Learning framework for telecommunications risk management".

Ighneiwa, I. & H. Mohamed (2017). "Bypass fraud detection: Artificial intelligence approach". arXiv preprint arXiv:1711.04627.

Jiménez Burgos, M. L. "El impacto del fraude en empresas de telecomunicaciones".

Kouam, A. J., A. C. Viana & A. Tchana (2021). "SIMBox bypass frauds in cellular networks: Strategies, evolution, detection, and future directions". IEEE Communications Surveys & Tutorials.

Menes Camejo, I., G. Arcos Medina & K. Gallegos Carrillo (2015). "Desempeño de algoritmos de minería en indicadores académicos: Árbol de Decisión y Regresión Logística". *Revista Cubana de Ciencias Informáticas,* 9(4): 104-117.

Pérez, A. M., *et al.* (2004). "Aplicaciones de aprendizaje no supervisado para la detección de patrones de fraude en telecomunicaciones". *Comunicaciones de Telefónica I+D*, (34): 161-180.

Subudhi, S. & S. Panigrahi (2016). "Use of fuzzy clustering and support vector machine for detecting fraud in mobile telecommunication networks". *International Journal of Security and Networks,* 11(1-2): 3-11.

Zhao, Q., K. Chen, T. Li, Y. Yang & X. Wang (2018). "Detecting telecommunication fraud by understanding the contents of a call". *Cybersecurity*, 1(1): 1-12.

Tesis de maestría

Getahun, A. (2020). Telecom Voice Traffic Termination Fraud Detection Using Ensemble Learning: The Case of Ethio Telecom, St. Mary's University.

Orbe Ordoñez, M. A. (2018). Propuesta metodológica de analítica de datos para estudio y análisis de tráfico en redes de telecomunicaciones, Quito, 2017.

Jornadas, congresos, conferencias

Tarmazakov, E. I. & D. S. Silnov (2018). Modern approaches to prevent fraud in mobile communications networks. IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), IEEE.