PORTELA, Gutierrez Portela; CÁRDENAS, Stefania Rodríguez; OSPINA, Laura Paola Patiño; AROS, Ludivia Hernandez. (2023) Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático. Cafi, v. 6 n. 1, p.77– 101. ISSN 2595-1750.
https://doi.org/10.23925/cafi.v6i1.58372

Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático[1]

Study of the prevention and detection of financial fraud through machine learning techniques

Recebido: 17/01/2023 - Aprovado: 20/03/2023 - Publicado: 01/04/2023
Processo de Avaliação: Double Blind Review

Fernando Gutierrez Portela[2]
Stefania Rodríguez Cárdenas[3]
Laura Paola Patiño Ospina[4]
Ludivia Hernandez Aros[5]

RESUMEN

Muchas organizaciones se ven afectadas actualmente por fraudes financieros convirtiéndose en una preocupación para el área financiera de cualquier entidad, ya que al materializarse perjudica directamente el patrimonio de cualquier empresa pública o privada. Para ello se han implementado técnicas supervisadas y no supervisadas que usan la inteligencia artificial para la prevención y detección temprana de estos fraudes y así minimizar riesgos en las operaciones financieras. Debido a lo anterior, el estudio analiza el uso de las técnicas supervisadas, su estado referencial por medio del análisis cienciométrico y bibliométrico, determinando la importancia de ellas para la prevención y detección de los fraudes financieros. A nivel metodológico es un estudio documental, exploratorio y analítico. Los resultados del estudio indican que las técnicas de aprendizaje automático supervisadas son las más precisas en el momento de aplicar los experimentos para la detección y prevención, logrando así resultados de efectividad superiores al 90% utilizando algoritmos como árbol de decisión, redes neuronales, Naive Bayes, Maquina de vectores de soportes, Bosque aleatorio y regresión logística, siendo notable en los resultados que los fraudes financieros mayormente analizados en estos estudios fueron falsificación de estados financieros, fraude de tarjetas de crédito, informes financieros fraudulentos y fraude de servicios financieros. Por otra parte, se resalta que el tema de investigación está en crecimiento gracias a que la detección de fraudes se está volviendo necesaria para las organizaciones y con mayor

[1] Estudio derivado de proyecto de investigación " ***Estudio de la detección del fraude empresarial en el sector de la economía social con el uso de técnicas de aprendizaje automático*** ", asociado al grupo PLANAUDI y AQUA, adscrito al Centro de Investigaciones del programa de Contaduría Pública e Ingeniería de sistemas de la Universidad Cooperativa de Colombia sede Ibagué.

[2] Candidato a Doctor en Ingeniería de la Universidad Autónoma de Bucaramanga. Magíster en Software Libre, Profesor de la facultad de Ingeniería de sistemas de la Universidad. Cooperativa de Colombia Sede Ibagué-Espinal. Integrante del Grupo de Investigación. AQUA de la UCC Ibagué-Espinal. Colombia. Correo: fernando.gutierrez@campusucc.edu.co.

[3] Estudiante de la Universidad Cooperativa de Colombia del programa de Contaduría de la sede Ibagué – Espinal. Colombia. Correo: stefania.rodriguezc@campusucc.edu.co.

[4] Estudiante de la Universidad Cooperativa de Colombia del programa de Contaduría de la sede Ibagué – Espinal. Colombia. Correo: laura.patinoo@campusucc.edu.co.

[5] Magister en auditoría y gestión empresarial de la universidad UNINI – puerto rico. Especialidad en revisoría fiscal y control de gestión de la universidad cooperativa de Colombia. Investigadora clasificada como asociado de Colciencias perteneciente a la facultad de contaduría pública universidad cooperativa de Colombia. Sede Ibagué, Colombia, grupo de investigación PLANAUDI. Correo: ludivia.hernandez@campusucc.edu.co.

PORTELA, Gutierrez Portela; CÁRDENAS, Stefania Rodríguez; OSPINA, Laura Paola Patiño; AROS, Ludivia Hernandez. (2023) Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático. Cafi, v. 6 n. 1, p.77– 101. ISSN 2595-1750.
https://doi.org/10.23925/cafi.v6i1.58372

relevancia para las instituciones financieras, por ser una de las mayores afectadas por este flagelo del fraude.
Palabras clave: Fraude financiero; Inteligencia artificial; Técnicas supervisadas; Detección de fraudes; Minería de datos.

ABSTRACT
Many organizations are currently affected by financial fraud, becoming a current concern for the financial area of any entity, since when it materializes, it directly harms the assets of any public or private company. To do this, and in response to this problem, supervised and unsupervised techniques have been implemented that use artificial intelligence for the prevention and early detection of these frauds and, thus, minimize risks in financial operations. Due to the above, the study analyzes the use of supervised techniques, their referential status through scientometric and bibliometric analysis, determining their importance for the prevention and detection of financial fraud. At the methodological level, it is a documentary, exploratory and analytical study. The results of the study indicate that supervised machine learning techniques are the most accurate when applying experiments for detection and prevention, thus achieving effectiveness results greater than 90% using algorithms such as decision trees, neural networks, Naive Bayes, Support Vector Machine, Random Forest and logistic regression, being notable in the results that the financial frauds, mostly analyzed in these studies, were falsification of financial statements, credit card fraud, fraudulent financial reports and service financial fraud. On the other hand, it is highlighted that the subject of research is growing thanks to the fact that fraud detection is becoming necessary for organizations and, with greater relevance, for financial institutions, as they are one of the most affected by this scourge of fraud.
Keywords: *Financial fraud; Artificial intelligence; Supervised techniques; Fraud detection; Data mining.*

1. INTRODUCCIÓN

Los fraudes financieros han sido uno de los mayores problemas económicos a nivel global que ocurre cuando un individuo incurre en afectaciones para la salud financiera de una compañía con el fin de que su poder haga más probable o factibles el delito financiero. Desde el surgimiento del caso Enron en 2001 y la debacle de WorldCom al año siguiente, el interés de los académicos por los escándalos empresariales ha crecido sustancialmente (Rosenblum, 2005). Enron Corporación, empresa estadounidense declarada en bancarrota culminando el año 2001, fue una compañía que incurrió en uno de los casos de fraudes financieros más conocidos e investigados en toda la historia, disfrazando las cifras reales de su información financiera para poder engañar a sus accionistas.

Como respuesta a los fraudes financieros presentados, actualmente se usa la Inteligencia Artificial (IA) y el aprendizaje automático como herramientas para brindar mayor facilidad a la hora de trabajar esto temas; sin embargo, el crecimiento de esta nueva

tecnología ha generado mayores riesgos de fraude que no han sido fáciles de detectar ya que cuando se desarrolla la herramienta, los timadores muchas veces ya han cambiado su forma de estafar.

La inteligencia artificial es el estudio de cómo construir o programar computadoras para hacer lo que la mente puede hacer (Boden, 1996), resulta demasiado útil en la vida cotidiana y ayuda de inmensa manera en algunos procesos, no tan solo contables o financieros, sino en varios campos de acción, como la ciencia, el entretenimiento, entre otros. Por lo anterior, este estudio analiza el uso de la inteligencia artificial implementada para la prevención y detección de los fraudes financieros.

Esta herramienta se ha convertido en un área avanzada de enseñar a las computadoras a imitar el cerebro humano y ha llevado el campo de la estadística a una disciplina amplia que produce teorías computacionales estadísticas fundamentales de los procesos de aprendizaje. El aprendizaje automático tiene una importancia inmensa en la bioinformática y la ciencia biomédica en general (Larrañaga et al, 2006).

Un algoritmo de aprendizaje automático es aquel que puede aprender de la experiencia con respecto a alguna clase de tareas y una medida de rendimiento (Mitchell, 1997). En general, hay dos tipos de esquemas de aprendizaje en el tema en cuestión: aprendizaje supervisado, en el que el resultado se ha etiquetado a priori o el alumno tiene algún conocimiento previo de los datos; y aprendizaje no supervisado en el que no se proporciona información previa al alumno sobre los datos o el resultado.

2. MARCO TEÓRICO

Mediante este trabajo de investigación se busca analizar las posibles causas de los fraudes económicos corporativas y de esa manera, plantear soluciones debido al impacto negativo generado a las entidades, propiciando la terminación de la entidad, como también generando un problema financiero en la sociedad.

2.1. FRAUDES ECONÓMICOS CORPORATIVOS Y SUS TIPOLOGÍAS

El fraude económico es bastante recurrente en todo el mundo. Lo que se denomina fraude difiere tanto de un momento a otro como de un lugar a otro. Los estudios de casos históricos de fraude parecen ser de especial mérito para investigar esto más a fondo (Cooper

et al., 2013). Puesto que la naturaleza de un fraude no se limita a un solo campo de acción; en los estudios sobre los fraudes financieros, se identifican tres elementos principales que son relevantes al momento de destapar el escándalo que podría llevar a la quiebra a una compañía.

El primer factor se identifica como el impulso obsesivo de cometer el delito sin importar los riesgos, teniendo pleno conocimiento de ellos. Sin embargo, la codicia, arrogancia y el optimismo obsesivo evidenciados en su proceder no le permite al actor del crimen identificar las consecuencias reales de estos actos. Como segundo factor se encuentra la complejidad de control de una compañía por su tamaño, lo que quiere decir que su complicado sistema -debido a un enorme crecimiento- genera dificultades a la hora de ejecutar el control interno de la empresa, facilitando que se cometan actos fraudulentos. Por último, se tiene como factor la presión competitiva, lo que quiere decir que el éxito de otra compañía genera recelo, que conlleva a probables prácticas sospechosas con el fin de aparentar una gloria no verídica, como el caso anteriormente mencionado.

El fraude ha provocado situaciones potencialmente impactantes en las empresas por lo que se estima que estas pierden 5% de los ingresos cada año. Por otro lado, los fraudes se clasifican como apropiación indebida de activos, corrupción, y fraude en los estados financieros, siendo la apropiación indebida de activos más común (James, 2014). El fraude es un problema complejo que se presenta constantemente en las empresas y que en la mayoría de los casos es poco detectable (Rozen, 2014).

Las personas que hacen parte de un fraude corporativo pueden ser miembros directivos que tiene el poder para tomar decisiones en la misma y aquellos empleados que están al alcance del objetivo como el área de tesorería, cartera y de compras, provocando actividades deshonestas, siendo capaz de llegar a un punto de desequilibrio económico, entre otras palabras pérdidas financieras (Linares Galan, 2022).

Por lo que a lo largo de los años y gracias al desarrollo de la tecnología, se han aplicado diversas soluciones con el objetivo de reducir los intentos de fraudes e imposibilitar su éxito en el mundo empresarial como en cualquier otro ámbito en el que se necesario. De acuerdo con lo anterior, el fraude financiero es una amenaza para las empresas a nivel operativo y económico (Fetzer, 1990).

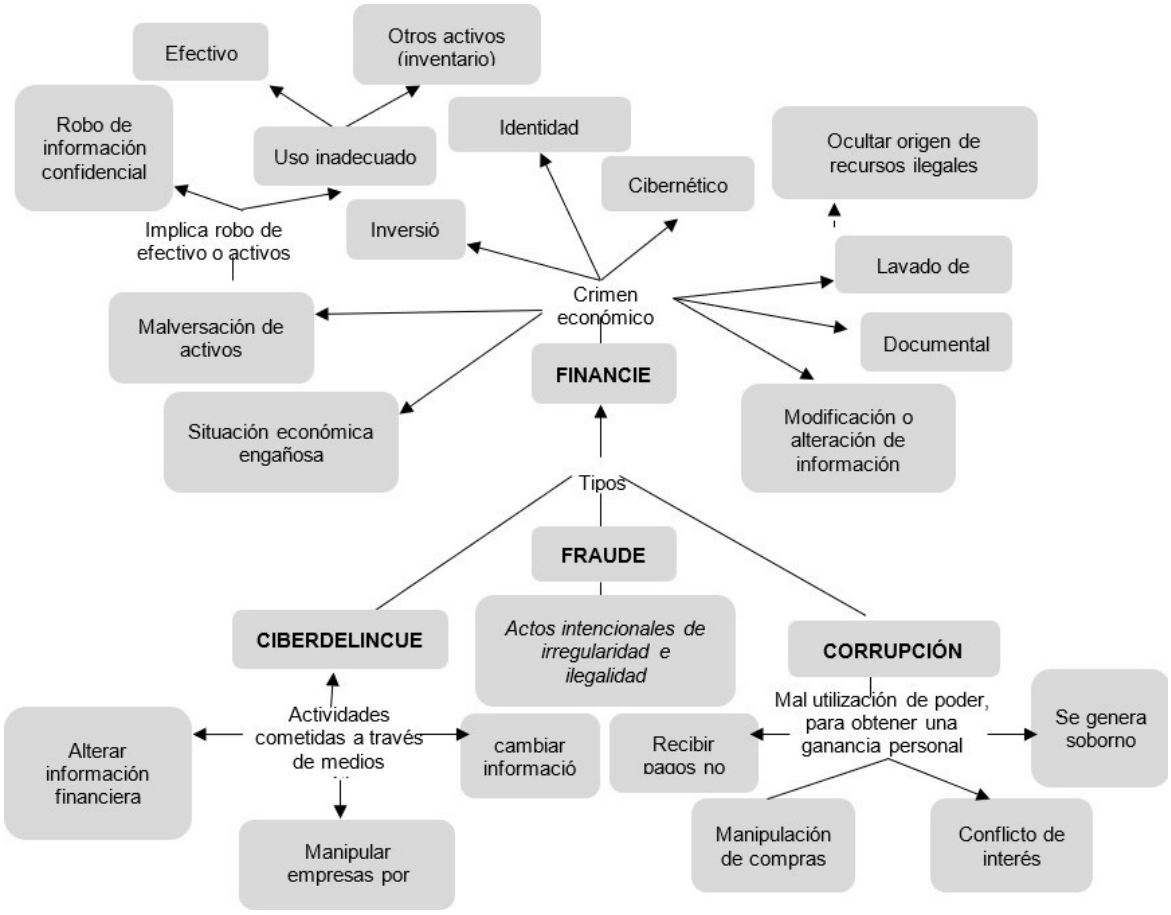
PORTELA, Gutierrez Portela; CÁRDENAS, Stefania Rodríguez; OSPINA, Laura Paola Patiño; AROS, Ludivia Hernandez. (2023) Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático. Cafi, v. 6 n. 1, p.77– 101. ISSN 2595-1750.
https://doi.org/10.23925/cafi.v6i1.58372

En cuanto a las tipologías del Crimen Financiero, se precisa como un acto de confianza aprovechados de actos engañosos, por lo cual se clasifican en: Malversación de activos, fraude financiero, corrupción y cibercrimen. Generalmente son perfiles diferentes que manejan actos ilícitos (KPMG, 2013).

De acuerdo con la encuesta a diferentes compañías en Colombia, se obtuvo un resultado preocupante con el tema de ejecución de los controles para detectar conductas relacionadas con el Fraude Financiero, debido a que, la persona evidencia que no tiene obstáculos para poder cumplir el objetivo de apropiación indebida de activos o entre otros casos ilícitos (KPMG, 2013).

A continuación, se relaciona en la figura 1 los tipos de fraudes financieros:

Figura 1 – Tipos de fraudes



Fuente: Los autores

La figura 1, identifica diversas categorías de fraudes que impactan en las finanzas de las compañías, tales como el fraude en malversación de activos, el fraude mediante el

reembolso indebido de gastos, la alteración de datos en los estados contables, lavado de activos, entre otros. Todos estos fraudes generan pérdida de recursos financieros de la empresa, reputación afectada, inciden negativamente en la moral y la confianza del personal que está a cargo de la dirección y el liderazgo. Ante esta problemática, se deben implementar medidas de prevención y detección de fraudes financieros, fomentar una cultura de ética empresarial y transparencia y actuar con rapidez si se descubren situaciones fraudulentas.

2.2. LOS FRAUDES FINANCIEROS Y LAS TÉCNICAS DE APRENDIZAJE AUTOMÁTICO APLICADAS EN SU DETECCIÓN Y PREVENCIÓN

La facultad de la computación a través de las últimas décadas ha crecido con rapidez en la ayuda de mejorar procesos en las empresas, y esto ha llevado a la transmutación de una nueva era generando una fortaleza hacia nuevos ingresos, nuevos procesos, nuevas inversiones con el fin de mejorar. Esto significa que el procedimiento tecnológico, como lo es la inteligencia artificial cuya finalidad es que las maquinas sean competentes para realizar actividades de la misma manera es que las realiza un humano, en la mayoría de los casos, se desarrolla a través de la ejecución de normas previamente programadas (Calvo et al., 2018).

A inicios de la década de los años 50’s, el término de las máquinas pensantes comenzaba a tomar relevancia en el campo de la cibernética y demás ramas relacionadas. Se le conoce como La conferencia de Dartmouth al primer proyecto de investigación donde por vez primera, se usó el concepto "Inteligencia Artificial" (IA). Para el verano de 1956 se conoció la fundación de la IA donde se entró a discusión sobre cualquier opinión o conjetura acerca del estudio o cualquier característica que está tecnología posee, cuando algunos veteranos de la computación militar temprana solicitaron a la Fundación Rockefeller una subvención de verano para financiar el taller que a su vez dio forma al campo. Existe una placa en Dartmouth College (universidad donde aconteció el suceso) que hace alusión a este importante evento y cita la conjetura central de su propuesta: que el comportamiento humano inteligente consistía en procesos que podían formalizarse y reproducirse en una máquina (McCarthy et al., 1955).

El aprendizaje automático para la detección y prevención es un modelo tecnológico que le permite a los programas aprender e implementar mejoras por sí mismos. Entre otros términos, es un tipo de inteligencia artificial, donde su eje principal se basa en la información

que se les da a los equipos tecnológicos para que, por sí mismos, reconozcan y sepan diferenciar por medio de patrones en algo en específico. La detención y prevención se da por medio de la utilización de algoritmos, para luego analizarlos, de acuerdo con los resultados, y de esa manera evidenciar anomalías.

La inteligencia artificial y el aprendizaje automático contienen gran cantidad de aplicaciones que fortalecen y mejoran la experiencia de todos los usuarios hasta permitir que las empresas luchen contra el fraude financiero, administrar y mitigar el riesgo (Ciobanu, 2019). El aprendizaje automático está en marcha en algunas empresas con el pleno objetivo de detección de fraudes en línea, la identificación de las anomalías sirve como base para luchar con patrones sospechosas. La función de este aprendizaje es abordar soluciones mediante el análisis de resultados de las bases de datos que posteriormente se utiliza para la toma de decisiones sobre riesgos. Existen técnicas de aprendizaje, las cuales se dividen por algoritmos "…capaces de aprender a partir de distintas y nuevas fuentes de información, construyendo algoritmos que mejoren de forma autónoma con la experiencia" (Calvo et al., 2018).

El aprendizaje supervisado es un conjunto de datos de entrenamiento etiquetados como resultado para aprender, problemas de clasificación y de regresión (Chirinos et al., 2021). Los cuales se pueden clasificar de acuerdo con la variable de su objetivo, por su similitud, el concepto de esta clasificación es colacionar cada petición nueva contra un modelo de cada clase usando alguna métrica, para después destinar la clase cuyo modelo sea más semejante a la petición (Langwagen Fripp, 2019). El algoritmo vecinos k-vecinos más cercanos, es el más practico ya que se utiliza como punto de referencia para clasificaciones más complejas, como redes neuronales artificiales y vectores de soporte (Chirinos et al., 2021).

Por su probabilidad, esta clasificación del aprendizaje supervisado utiliza el fundamento de máxima apariencia u acercamiento para determinar decisiones (Langwagen Fripp, 2019), el algoritmo Naïve Bayes es un enfoque estadístico basado en la teoría bayesiana, que selecciona las decisiones basadas en la mayor probabilidad, junto a la regresión logística que encuentra el parámetro que mejor se ajuste para estimar la probabilidad de una respuesta binaria para una o más características, son los algoritmos más usados para para la clasificación de probabilidad (Chirinos et al., 2021).

Las Máquinas de Vectores de Soporte son un tipo de algoritmo de aprendizaje automático supervisado que se utiliza principalmente para la clasificación y regresión. Su objetivo es encontrar un hiperplano en un espacio multidimensional que mejor separe las clases de datos. El algoritmo busca maximizar el margen entre las clases, es decir, la distancia entre los puntos más cercanos de las diferentes clases (Langwagen Fripp, 2019), esta clasificación tiene varios algoritmos que se caracterizan organizadas con una estructura jerárquica para decidir qué se ajusta a cumplir con sus condiciones de raíz a hoja (Chirinos et al., 2021).

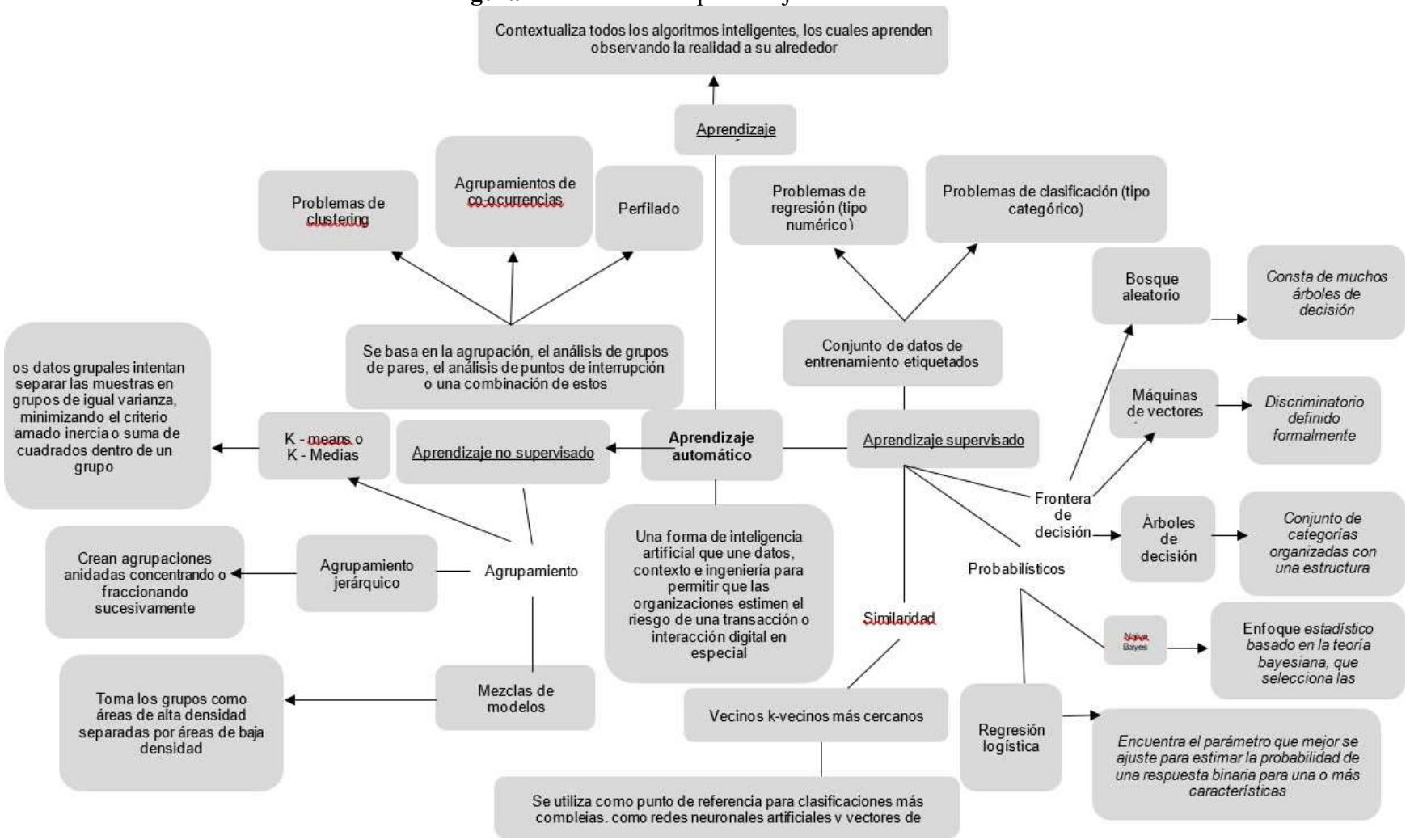
El aprendizaje no supervisado se basa en la agrupación, el análisis de grupos de pares y el análisis de puntos de interrupción o una combinación de estos (Chirinos et al., 2021), se clasifican en problemas agrupación los cuales encuentran una organización o un modelo en una colección de datos no categorizados, por asociación constituye entre objetos de datos dentro de grandes bases de datos (González, 2020).

Los algoritmos más comunes del aprendizaje no supervisado por agrupamiento, k-means o K-Medias son datos grupales que intentan separar las muestras en n grupos de igual varianza, minimizando el criterio llamado inercia o suma de cuadrados dentro de un grupo, él algoritmo por agrupamiento Jerárquico estos crean agrupaciones anidadas fraccionando sucesivamente y mezclas de modelos gaussianos toma los grupos como áreas de alta densidad separadas por áreas de baja densidad (Chirinos et al., 2021).

La técnica de aprendizaje por refuerzo, la cual contextualiza todos los algoritmos inteligentes, los cuales aprenden observando la realidad a su alrededor, La recompensa puede tener un cierto retardo y una fuerte carga de ensayo y error (Chirinos et al., 2021).

PORTELA, Gutierrez Portela; CÁRDENAS, Stefania Rodríguez; OSPINA, Laura Paola Patiño; AROS, Ludivia Hernandez. (2023) Estudio de la prevención y detección de fraudes financieros a través de técnicas de aprendizaje automático. Cafi, v. 6 n. 1, p. 76 – 97. ISSN 2595-1750.
https://doi.org/10.23925/cafi.v6i1.58372

Figura 2 - Técnicas de aprendizaje automático



Fuente: Los autores

En la figura 2, se resume las diferentes clases de aprendizaje automático como disciplina de la inteligencia artificial que engloba diferentes enfoques: aprendizaje supervisado que entrena al modelo con ejemplos etiquetados para hacer predicciones precisas; no supervisado que busca descubrir patrones y estructuras en datos sin etiquetas, y por refuerzo el cual se basa en la interacción con un entorno para maximizar las recompensas. Ahora bien, en el contexto de la detección de fraude financiero, se emplean diversas técnicas de aprendizaje automático como la detección de anomalías que se enfoca en identificar patrones sospechosos, el aprendizaje supervisado se usa para clasificar transacciones como fraudulentas o legítimas, y los modelos de Deep Learning y el procesamiento del lenguaje natural son aplicados para abordar fraudes más sofisticados.

En el ámbito de la detección del fraude financiero, se utilizan métodos como la detección de anomalías, el aprendizaje supervisado, las redes neuronales y el procesamiento del lenguaje natural para identificar patrones sospechosos y proteger los activos de las empresas. Estas herramientas avanzadas contribuyen a prevenir fraudes y mantener la confianza en el sector financiero.

3. METODOLOGÍA

El presente documento extiende un estudio de tipo documental, exploratorio y analítico, investigada por Hernández Sampieri et al. (2018), recordando que la primera se soporta de una revisión de archivos u notas académicas de tipo investigativo, en el que se referirán a diversos autores que refuten el hoy tema de investigación; la sesión exploratoria, define una serie de hallazgos que permitan, o bien, entender el fenómeno puesto a la investigación, al caso, o evaluar y recopilar información dispersa que se encuentra dentro de fuentes oficiales, como confiables. Por último, encimar un aspecto analítico, el cual mida los datos e información recopilada y le dé estructura al documento de tenencia, es decir, se toman los archivos recopilados para dar un veredicto de lo agrupado.

De forma tal que las acciones cualitativas para esta investigación superan en calidad de respuesta los datos cuantitativos, de igual importancia para calcular la magnitud del fenómeno, sin embargo, su haber es mucho más especulativo; el fraude tiene características complejas y en sus diversas formas es difícil asignarles números, más que por cantidad. El

estudio inspecciona, según los preceptos de Hernández Sampieri et al. (2018), las técnicas en mención, supuestos de valor cualitativo, para implementar en la prevención, detección y detención de los fraudes financieros, fruto del presente estudio y del cual se basa el concepto cuantitativo, netamente para revisar su magnitud.

Por cuanto, el último avance de los tres tipos documental, exploratorio y analítico prevé el sistema analítico propuesto por Hurtado de Barrera (2012), una inspección a los referentes académicos, autores, datos e información en general. Se tiene para ello el sistema analítico según Hurtado de Barrera, reflectando los mejores conceptos que se acomoden al espectro de la investigación propia, resaltando cuáles son más relevantes e impactantes a los resultados esperados, lo que es igual a dejar un espacio para determinar características y conceptos claves.

Para la investigación se tuvo en cuenta las siguientes fases:

Fase 1. Se estableció la ecuación de búsqueda, como proceso de ajuste a la información, señalando el conjunto de términos y características útiles de los referentes bibliográficos y dotándoles de personalidad a la investigación, con el fin de agilizar el proceso de búsqueda.

Fase 2. Se inició con una exploración bibliográfica sobre diversas fuentes y autores oficiales, como confiables, aplicando un análisis cienciométrico, cuyo principal ejercicio era el de medir el impacto, establecer conjuntos de referencias académicas, comprensión de citas, origen de indicadores de resultados y el mapeo hipótesis, investigaciones, referentes, artículos, periódicos, etc.

La ecuación aplicada para poder llegar al resultado según la base de datos de Scopus (https://www-scopus-com.bbibliograficas.ucc.edu.co/search/form.uri?display=basic#basic):

((TITLE-ABS-KEY ("machine learning")) AND (("supervised Learning")) AND ("financial fraud") AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2007))

Tomada el 2 de mayo a las 8:00 pm.

Fase 3. Se diseñó y analizó un cuadro experimental de autores y referencias determinadas en la segunda fase, con la idea de parametrizar y comparar la información, dejando claro los hechos irrefutables y cuáles son algunas de las inconsistencias más visibles.

Fase 4. Se inspeccionó y analizó los estudios de casos en organizaciones con situaciones exitosas en la detección y prevención de fraudes financieros, a fin de traerlas a colocación y señalar las técnicas utilizadas que lograron el mencionado éxito.

Fase 5. Se analizaron las técnicas supervisadas más implementadas con éxito, haciendo un paralelismo entre ellas, señalando factores comunes, al igual que su convergencia entre dos o más de las técnicas.

Finalmente, el alcance de la investigación, según Molina (2017), se ve direccionado a la especulación de la eficiencia de las técnicas para la prevención y detección de los fraudes financieros, dado que la realidad de su aplicación es otro asunto arraigado a la ética profesional, por ejemplo, u otros que involucran diversos agentes que nublan un resultado claro de su eficacia, siempre y cuando se quiera cometer un fraude, es posible si existen los canales para ello, en cuento al estudio, permitirá resaltar las mejores prácticas para reducirlo, mas no para detenerlo.

Se explicó los antecedentes e interpretó los hallazgos de técnicas aplicadas, la solución a las variables e hipótesis del fraude financiero y su comportamiento con las mismas, con el fin de aprobar un nuevo esquema actualizado de los avances de dichas técnicas al combatir el fraude y como siguen siendo o no eficientes, así mismo, finalizar algunas recomendaciones sobre sus fallos en la gestión.

4. RESULTADOS

4.1. FRAUDES FINANCIEROS Y TÉCNICAS DE APRENDIZAJE SUPERVISADAS HAN SIDO USADAS EN LA DETECCIÓN Y PREVENCIÓN TEMPRANA POR LAS ORGANIZACIONES

El impresionante avance tecnológico en la humanidad ha permitido que cada importante industria implemente nuevos sistemas para agilizar sus procesos productivos en toda clase de aspectos relacionados con el desarrollo económico, social, productivo, entre otros. Uno de estos avances es el uso de algoritmos y técnicas de aprendizaje automático que permitan detectar al instante eventos de fraude que quieran cometerse, todo esto aplicado en el área financiera de entidades empresariales. Un número de importantes profesionales de la

materia han utilizado estas herramientas con el objetivo de minimizar acciones fraudulentas en su contra y es que este sistema ha evolucionado de tal manera que las opciones de detección de fraude varían en función de la necesidad específica.

Gracias a los algoritmos y las técnicas de aprendizaje automático, el índice de actividades ilícitas ha disminuido considerablemente puesto que previamente, cuando el avance tecnológico no había tenido tanta magnitud era mucho más fácil proceder al engaño y estrategias ilegales para el beneficio propio. Sin embargo, es claro que siempre habrá quienes encuentren fallas de seguridad y logren llevar a cabo su cometido y es precisamente por esto que esta clase de medidas de seguridad funciona de manera tan eficiente, puesto que, como su nombre lo indica, el sistema aprende por su cuenta tomando una gran cantidad compilada de datos con el objetivo de aprender patrones, notar fallas potenciales y, sobre todo, proteger valiosos recursos financieros.

https://doi.org/10.23925/cafi.v6i1.58372

Tabla 1- Cuadro experimental

Articulo	Autor	Técnica	Fraude Detectado	Conjunto De Datos	Metricas-Accuracy
Técnicas de clasificación de minería de datos (DM) en la detección de empresas que emiten estados financieros fraudulentos (FFS)	Efstathios Kirkos and Charalambos Spathis and Yannis Manolopoulos-May 2007	Árbol de decisión	Falsificación de estados financieros	Nuestra muestra contiene datos de 76 empresas manufactureras griegas (sin incluir empresas financieras). Los auditores revisaron todas las empresas de la muestra. Para 38 de estas, fue publicada indicación o prueba de intervención en emisión de FFS.	Clasificó correctamente todos los casos sin fraude (100 %) y 35 de los 38 casos de fraude (92%)
		Redes neuronales			La red logró clasificar correctamente todos los casos, logrando así un rendimiento del 100%.
		Naïve Bayes			La red clasificó correctamente 72 casos (rendimiento 95%). En particular, clasificó correctamente 37 casos de fraude (97%) y 35 casos de no fraude (92%).
Minería de datos para el fraude con tarjetas de crédito: un estudio comparativo/Data mining for credit card fraud: A comparative study	Bhattacharyya, Siddhartha et al. (2011)	Máquinas de vectores de soporte (MVS)	Fraude de tarjetas de crédito	Este conjunto de datos tiene 13 meses, de enero de 2006 a enero de 2007, de alrededor de 50 millones (49,858,600 transacciones) transacciones con tarjetas de crédito en alrededor de un millón (1.167.757 tarjetas de crédito) tarjetas de crédito de un solo país A los fines de este estudio, llamamos a este conjunto de datos de todas transacciones, conjunto de datos U	0,938
		Bosque aleatorio			0,947
		Regresión logística			0,962
Detección de Fraude en TC Mediante Técnicas de Minería de Datos	González Martínez, E. F., Romero, G. R., Ortiz Rico, A. F., & Cruz Castro, D. L. (2018, junio)	Bosque aleatorio	Fraude en Tarjetas de Crédito (TC)	Los conjuntos de datos contienen transacciones realizadas con tarjetas de crédito en septiembre de 2013 por titulares de tarjetas europeos. Este conjunto de datos presenta las transacciones que ocurrieron en dos días, donde tenemos 492 fraudes de 284,807 transacciones.	99,96%
		Naïve Bayes			97,81%
		MVS			99,94%
Un análisis sobre la detección de fraude en los estados financieros para las empresas chinas que cotizan en bolsa que utilizan el aprendizaje profundo	Wu Xiuguo,Du Shengyong	Regresión logística	Fraude en estados financieros	Muestra contenía datos de 1068 empresas chinas distintas que cotizan en la Bolsa de Valores de Shenzhen entre 2016 y 2020	0.81
		Bosque aleatorio			0.87
		MVS			0.83

Articulo	Autor	Técnica	Fraude Detectado	Conjunto De Datos	Metricas-Accuracy
Informes anuales corporativos mineros para la detección inteligente del fraude en los estados financieros: un estudio comparativo de los métodos de aprendizaje automático	Petr Hajekun, Roberto Henriques	Regresion logística	Informes financieros fraudulentos	311 empresas públicas involucradas en presuntos casos de informes financieros fraudulentos durante el período 2005-2015 y, por lo tanto, se utilizó como muestra un conjunto de 311 informes anuales	74.53 ± 2.96
		Naïve Bayes			57.83 ± 3.61
		MVS			77.95 ± 3.05
Detección de transacciones financieras fraudulentas mediante máquina Aprendizaje	Mosa MM Megdad, Bassem S. Abu-Nasser y Samy S. Abu-Naser	Árbol de decisión	Fraude de servicios financieros	El conjunto de datos consta de 6362620 registros con 10 características. El valor medio de todas las transacciones es 144972 USD, mientras que la transacción más grande registrada en este conjunto de datos asciende a 1991430 USD	Conjunto de datos no balanceado: 99.7% Conjunto de datos balanceado:99.9%
		Regresión logística			Conjunto de datos no balanceado: 99.1% Conjunto de datos balanceado:96.6%

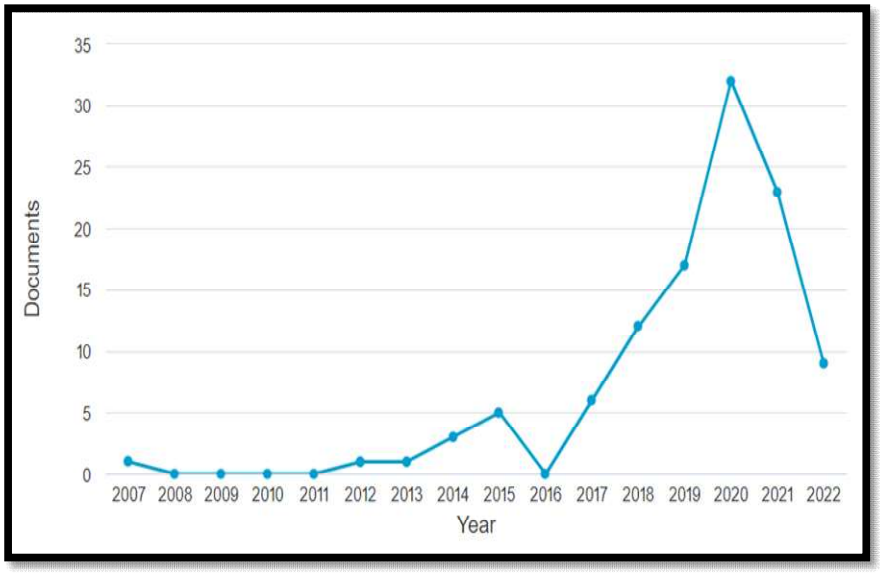
Fuente: Los autores.

En la tabla 1 se abordan diferentes autores que han usado en su análisis de fraude conjuntos de datos de la bolsa de valores de diferentes países o datos propios de sus empresas, con los resultados de precisión y otras métricas que evalúan que tan efectivo es la técnica usada, donde se observa que las investigaciones se centran los fraudes realizados con tarjeta de crédito y fraude en informes financieros fraudulentos.

4.2. TÉCNICAS DE APRENDIZAJE SUPERVISADAS PARA LA PREVENCIÓN Y DETECCIÓN DE LOS FRAUDES FINANCIEROS POR MEDIO DE LA REVISIÓN BIBLIOMÉTRICA

En esta revisión bibliométrica, después de aplicar la ecuación con las palabras claves de la investigación, se obtuvo un total de resultado de 110 documentos en la plataforma Scopus.

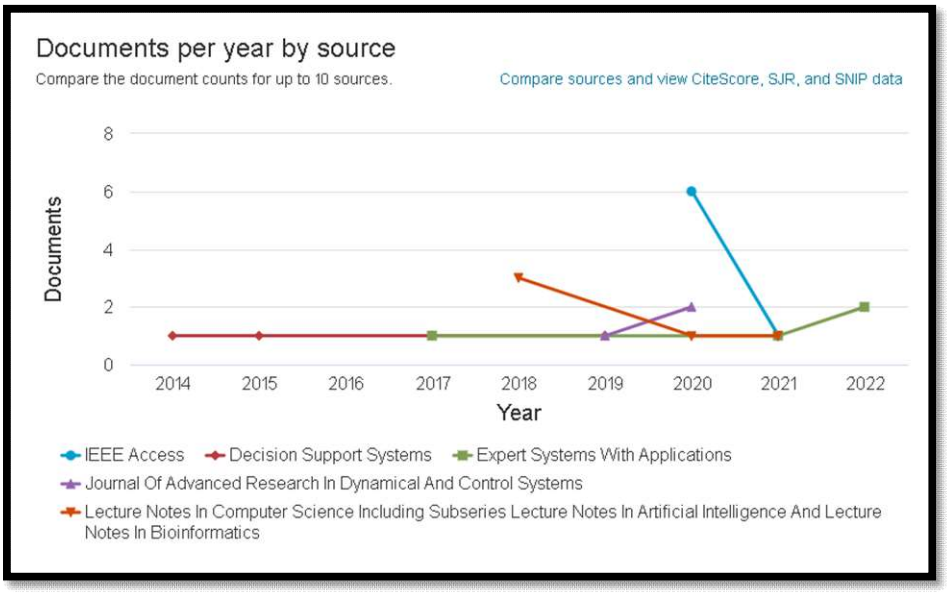
Figura 3 – Análisis de la cantidad de documentos por año publicados Scopus



Fuente: Figura extraída de la base de datos de SCOPUS.

La figura 3, muestra el análisis de la cantidad de artículos publicados cada año, en el que se observa que, en el período de 2007 a 2013 se dio inicio en las investigaciones del tema en curso. A partir del 2016, la investigación del uso de la IA para la detección de fraudes financieros ha aumentado, siendo un tema de gran relevancia en la actualidad.

Figura 4 – Documentos por año



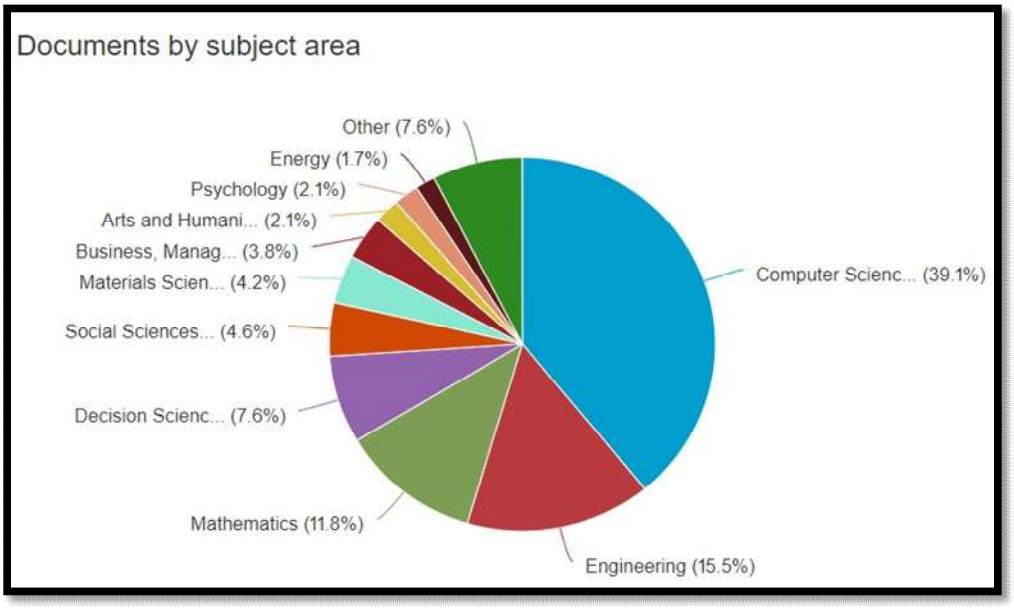
Fuente: Figura extraída de la base de datos de SCOPUS.

La figura 4 presenta las diferentes revistas que sobresalen en el tema de estudio como IEE Acces que ha venido publicando a partir de 2020, Journal of Advance Researh in Dynamical and Control Systems a partir del 2019 y notas en diferentes revistas que inicio desde el 2018, pero que ha venido decreciendo en sus investigaciones, al igual que Decisión Support Systems. Sobresale la revista de Expert Systems with applications viene analizando el tema desde el 2017 y que continúa publicando este tema de gran interés para la comunidad científica y el sector empresarial que son los directamente afectados en caso de materializarse un fraude.

El análisis de los artículos filtrados en función de su fuente de publicación. y su investigación se centra en el acceso al Instituto de Ingenieros Eléctricos y Electrónicos (IEEE), cursos y lecturas sobre ciencia, computación y series sobre inteligencia artificial y bioinformática

a. Los autores más representativos durante la investigación de las técnicas supervisadas, y compara el número de citas que le dieron a cada autor, entre las que destacan: Delen, D, este autor solo ha compilado 2 citas en los años 2014 a 2015, este es claramente un tema prometedor debido a la gran dificultad, y todos los demás autores destacados como HE, J, Kuzey, C, Maciejewski, R, entre otros, tienen el mismo número de veces que han sido citados en sus estudios.
b. Documento por área de estudio:

Figura 5 – Documentos por área temática



Fuente: Figura extraída de la base de datos de SCOPUS.

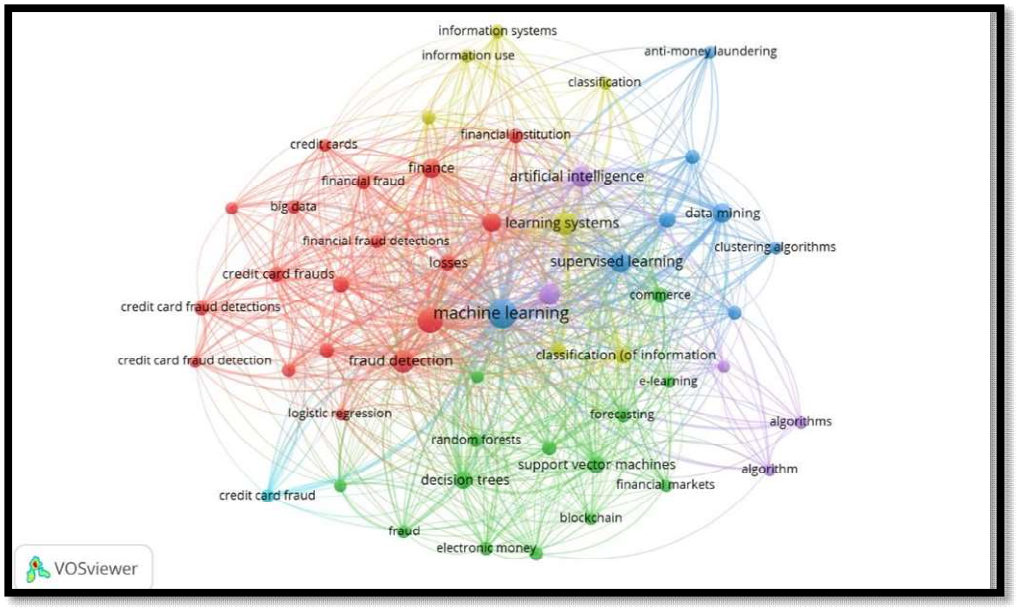
La figura 5 revela que los autores centraron su campo de investigación principalmente en el tema de Ciencias de la computación, ya que se evidenció el mayor número de publicaciones, con un total de 93 documentos en la base de datos Scopus, que representa el 39,1% de todos los artículos en una búsqueda, seguidamente de la ingeniería, tema importante porque es una rama muy enfocada de la inteligencia artificial que puede identificar técnicas de aprendizaje supervisado, en este caso el número total de documentos es 37, que es el 15,5% del análisis.

c. En el caso de analizar las publicaciones por ciudad o territorio, el sitio principal es Estados Unidos con un total de 19 documentos, en segundo lugar, se encuentra China con un total de 16 documentos en la encuesta y seguimiento. Le siguen países como India., Reino Unido, Australia; Con menor participación se encuentra países como Canadá, Malasia, Turquía, Arabia Saudita y Brasil.

En cuanto al tipo de documentos, de los 110 artículos analizados más utilizados para las investigaciones, representando el 50% de la encuesta, y en el segundo se pueden mencionar los artículos más populares. El tipo de documento es un documento de sesión, con un 31,8% de participación, y no se utilizan otro tipo de documentos, pero aún menos importantes, como revistas, libros, editoriales, etc.

d. Correlación por palabras clave

Figura 6 - Correlación por palabras clave



Fuente: Figura extraída de VOSVIEWER

La figura 6 presenta los documentos relacionados en la herramienta de VOSviewer, permitieron que los artículos fueron compilados en base a palabras clave, resúmenes y títulos, con un total de 53 palabras más relevantes y de acuerdo con el tema de investigación. De esta forma, se estructuró la correlación de las palabras clave sugeridas por el autor y por la base de datos Scopus, de la cual se extrajeron las palabras clave para poder determinar el enfoque de la investigación.

Se identificaron seis grupos de clúster a partir de la red bibliográfica presentada anteriormente, en el primer clúster color rojo se evidenciaron 18 palabras claves que encaminaron la investigación al área de inteligencia artificial y el fraude financiero, en el segundo clúster color verde se hallaron 14 palabras, el tercer clúster color azul con 8 palabras, el cuarto clúster color amarillo con 7 palabras, el quinto clúster color morado con 5 palabras claves y finalmente el sexto clúster color azul cielo con una palabra clave.

4.3. CASOS EXITOSOS FRENTE AL USO DE LAS TÉCNICAS DE APRENDIZAJE SUPERVISADAS PARA LA DETECCIÓN Y PREVENCIÓN DE FRAUDES FINANCIEROS

En los últimos años el machine learning (aprendizaje automático) ha sido una herramienta que ha superado a grandes rasgos lo que se esperaba de ella dentro de las organizaciones muchas de las empresas que afirman que han utilizado la tecnología de la inteligencia artificial para poder aumentar la productividad de sus empresas. Sin embargo, en esta parte se van a dar a conocer algunos casos de entidades que han utilizado técnicas de aprendizaje supervisadas para ser óptimos en el descubrimiento y prevención de fraudes financieros.

Estudios analizan las técnicas de minería de datos para la detección de Estados Financieros Fraudulentos (Kirkos et al., 2007) quienes decidieron aplicar tres técnicas de aprendizaje supervisadas con el fin de descubrir empresas que emitían estados financieros fraudulentos, usaron tres técnicas de la clasificación de minería de datos: árboles de decisión, las redes neuronales y las redes de creencias bayesianas. Son tres tipos de experimento que fueron comparados para medir su precisión predictiva, con una muestra de 76 empresas manufactureras de Grecia. Cada experimento fue evaluado dos veces y todos detectaron diferentes resultados en ambos estudios.

El caso con más éxito fue el árbol de decisiones en ambas partidas, que logró calcular de manera adecuada 73 casos, dando así una equivalencia del 96%; más concretamente, el árbol de decisiones clasificó correctamente todos los casos sin fraude (100 %) y 35 de los 38 casos de fraude (92 %). Comparado con el modelo de redes neuronales con ayuda de un software, que aportó en la construcción de una red de alimentación directa de perceptrones de múltiples capas. Después de varios diseños, se seleccionó una sola red que logró un rendimiento del 100%. Desafortunadamente, el software no proporcionó una interfaz transparente para los pesos sinápticos de las conexiones y, por lo tanto, no fue posible estimar la importancia de cada variable de entrada. En este escenario ambas técnicas fueron exitosas, la primera de manera más precisa y transparente para el análisis de los resultados de empresas que dan información financiera fraudulenta, para saber si el fraude gerencial es practicado a grandes rasgos (Kirkos et al., 2007).

Se analiza en la literatura casos de estudio en fraudes de tarjetas de crédito de tarjetas de crédito, aplicando técnicas supervisadas: máquinas de vectores de soporte (SVM) y

bosques aleatorios. Estos dos experimentos junto con la regresión logística también hacen parta de la minería de datos. Considerado que con el avance tecnológico que se ha obtenido de manera global, así mismo hay organizaciones dedicadas a los fraudes en línea que día tras día van mejorando sus técnicas para cometer los fraudes, es por esto que en este experimento combinan tres técnicas: la SVM, que es fuerte en fundamento teórico y aplicación exitosa para en una variantes de problemas; los árboles de decisión que, aunque es conocida por la facilidad de su uso, podría llegar a presentar problema de inestabilidad y confiabilidad; y por último la regresión logística que proporciona una base de utilidad para comparar el rendimiento de las técnicas más nuevas.

En la indagación se observó otros estudios donde se establece un modelo predictivo que usa la minería de datos aplicada en una fiducia para detectar el riesgo de liquidez. Así como lo bancos las entidades fiduciarias a raíz de ver la información de sus Estados financieros y los riesgos de liquidez que van ligados a la fiducia, siendo captadoras del recurso de inversión (León Sánchez, 2015).

Adicionalmente es en este caso de éxito en que se analiza la detección de riesgos operativos, donde por medio de un perfil, se desarrolló un estudio que se dividió en tres fases para poder identificar factores de riesgo operacional que afectan el correcto desarrollo de las firmas. En el que en la tercera fase se obtuvo datos que dieron paso a un buen análisis de los Estados Financieros implementando el árbol de decisiones conocida como técnicas supervisadas de la minería de datos.

En la aplicación a riesgo de crédito su técnica se centró en usar el método de las redes neuronales, este modelo fue realizado en dos bases: en la primera permitió hacer un grupo de datos homogéneos y desechó datos aislados que no eran tan relevantes para la obtención de resultados, de esta manera se hace un proceso de reclasificación de etiquetas para obtener un resultado mucho más preciso; lo anterior, comparando con estudios de datos de crédito de un banco local en China, donde se observó que, al seleccionar un punto de corte adecuado, se logra una mayor precisión en la clasificación de créditos buenos y malos (León Sánchez, 2015).

Este estudio (León Sánchez, 2015) se usó información abierta publicada de la Superintendencia Financiera de Colombia, el Banco de la República, donde se cuenta con una amplia calidad de registro y variables, el análisis comparativo de este experimento da

como resultado que, del comparar tres tipos de redes neuronales con otro modelo, se concluye que es el más utilizado para poder predecir los riesgos de liquidez en todo el sector financiero colombiano. Mientras que la técnica de árboles de decisiones se utilizó más para poder descubrir reglas y patrones que van más relacionadas al comportamiento, que se convierten en parte fundamente para la creación de un modelo para la predicción de en riesgo de liquidez y todo lo que implica, así para poder tener una toma de decisiones más más precisa respecto a este tipo de problema financiero.

Como ya se ha planteado, los fraudes financieros son una problemática que va avanzando día a día a medida que va aumentando la evolución tecnológica que se está obteniendo de manera global, ya que, este tipo de herramientas facilita el acceso a nuevas técnicas que puedan aplicar las organizaciones o personas dedicadas a esta actividad.

Esto es posible gracias a un conjunto de minería de datos donde se conocen dos tipos de técnicas: las supervisadas y las no supervisadas, en este caso se tiene un enfoque más preciso sobre las supervisadas, ya que por su exactitud son más confiables y efectivas en el momento de dar resultados para poder dar un buen análisis en cada de los estudios en los que se decide usar este tipo de técnicas.

5. CONCLUSIONES

Este estudio se analizó que, al implementar la inteligencia artificial en una compañía se puede evitar pérdidas económicas importantes, producto de un fraude financiero. Por otra parte, el uso de la innovación, acompañada de la minería de datos y las técnicas de aprendizaje automático, son una gran ayuda para la optimización de procesos y la minimización de riesgos.

En la investigación llevada a cabo en este estudio, se observó que las técnicas supervisadas, tienen mayor aplicación en los estudios analizados, donde la mayor parte de las técnicas son aplicadas utilizando modelos que muestran resultados de precisión diferenciales. En la detección de los fraudes, los experimentos utilizan técnicas en el campo del aprendizaje automático y la inteligencia artificial para construir modelos predictivos como: árbol de decisión, redes neuronales y máquinas de vectores de soportes, entre otras.

Luego de hacer el análisis cienciométrico y bibliométrico, se observó que el tema de las técnicas de aprendizaje automático en la detección del fraude, en la actualidad tienen gran acogida y se resalta la importancia de seguir indagando en el tema, debido a que existen variedad de técnicas como supervisadas, no supervisadas, semi-aprendizaje, por refuerzo, lo cual hace atractivo este tema en líneas de investigación futura, donde la inteligencia artificial interviene para ayudar a minimizar riesgos, lo cual hace que sea un factor clave en las organizaciones para la toma de decisiones, ya que funcionan como estrategia para el mejoramiento de procesos.

Frente a los casos de éxito, se analizó que el riesgo de liquidez de una cartera colectiva en Colombia, procesando un base de datos con un volumen considerable de información, en la León Sánchez, evaluó los resultados obtenidos y presentó el riesgo de liquidez de la fiducia aplicando directamente el experimento; aun así, en este estudio los modelos sistemáticos pueden aplicarse en diferentes tipologías de negocios, lo cual hace que la aplicación de la minería de datos sea efectiva.

Para finalizar, se concluye que es muy importante que se implemente nuevos avances tecnológicos, usando la inteligencia artificial al interior de las organizaciones, y lograr que día a día se esté en continuo mejoramiento en todos los procesos; lo anterior, debido a la amplia posibilidad de un fraude, y el alto riesgo operativo y económico por falta de nuevos controles, que den respuesta inmediata y hasta predictiva a nuevas modalidades de fraude.

Para lo cual se hace necesario que se analicen nuevos modelos predictivos con sus técnicas de aprendizaje automático en casos de fraude, que otorguen herramientas de bloqueo ante un eminente fraude. Además, se considera que las grandes compañías de auditoría deben ir de la mano con la academia para que la información que estas compañías manejan en sus auditorías sirvan de base para los experimentos, evitando el uso de datos sintéticos.

REFERENCIAS

Boden, M. (1996). Artificial Intelligence. Elsevier. Obtenido de books.google.es: https://books.google.es/books?id=_ixmRlL9jcIC&printsec=frontcover&hl=es#v=onepage&q&f=false

Calvo, J., Guzmán, M., & Daniel, R. (2018). Machine learning, una pieza clave en la transformación de los modelos de negocio. Management solutions Making things

happen. Obtenido de managementsolutions.com: https://www.managementsolutions.com/sites/default/files/publicaciones/esp/machine-learning.pdf

Chirinos, Y., Godínez López, R., & Ramírez García, A. (2021). Tendencias Investigación Universitaria Vol. IX.Cáp IV. Obtenido de researchgate.net: https://www.researchgate.net/publication/349607908_Tendencias_Investigacion_Universitaria_Vol_IX

Ciobanu, M. (2019). The rise of machine learning and artificial intelligence in fraud detection. Obtenido de thepaypers.com: https://thepaypers.com/expert-opinion/the-rise-of-machine-learning-and-artificial-intelligence-in-fraud-detection/779255

Cooper, D., Dacin, T., & Palmer, D. (2013). Fraud in accounting, organizations and society: Extending the boundaries of research. Accounting, Organizations and Society, 38, 440–457. Obtenido de webofscience.com: https://www.webofscience.com/wos/woscc/full-record/WOS:000329013200003?SID=7F2IsIgNMPe7vfo858n

Fetzer, J. H. (1990) Artificial Intelligence: Its Scope and Limits. Kluwer Academic Publishers. Recuperado de: https://books.google.com.co/books?id=77SLBQAAQBAJ&printsec=frontcover&hl=es#v=onepage&q&f=false

González, L. (18 de agosto de 2020). Aprendizaje no Supervisado. Obtenido de aprendeia.com: https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/

Hernández Sampieri, R., & Torres, C. (2018). Metodología de la investigación (Vol. 4). México DF: McGraw-Hill Interamericana.

Hurtado de Barrera, J. (2012). El Proyecto de Investigación. Qué es investigar. El "quiénes" de la investigación (Séptima Edición). Caracas, Venezuela: Ediciones: Quirón (SYPAL).

James, R. (2014). Report to the Nations on Occupational Fraud and Abuse: 2014 Global Fraud Study, 79: 1-79.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. Expert systems with applications, 32(4), 995-1003

KPMG. (2013). Encuesta de fraude en Colombia. (KPMG, Ed.) Bogotá: KPMG Colombia.

Langwagen Fripp, L. (octubre de 2019). Aplicación de aprendizaje automático a la detección de fraude en tarjetas de crédito. Obtenido de colibri.udelar.edu.uy (Universidad de la república Uruguay): https://www.colibri.udelar.edu.uy/jspui/bitstream/20.500.12008/23163/1/Lan19.pdf

Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., ... & Robles, V. (2006). Machine learning in bioinformatics. Briefings in bioinformatics, 7(1), 86-112.

León Sánchez, D. P. (2015). Modelo predictivo para riesgo de liquidez de una entidad fiduciaria usando minería de datos (Doctoral dissertation).

Linares Galán, J. E. (2022). Control interno en la prevención y detección de fraude corporativo [Diapositivas]. Obtenido de jabrveriana.edu.co: https://www.jabrveriana.edu.co/personales/hbermude/Audire/jelg2.pdf

McCarthy, J., Minsky, M., Rochester, N. y Shannon, C. (1955). Una propuesta para el proyecto de investigación de verano de Dartmouth sobre inteligencia artificial. Consultado el 15 de octubre de 2022. Archivado desde el original en http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

Mitchell, T. (1997). Machine Learning. McGraw-Hill.

Molina, R. (2017). ¿Formación para la investigación o investigación formativa? La investigación y la formación como pilar común de desarrollo. Revista Boletín Redipe, 6(1), 84-89.

Rosenblum, H. (01 de enero de 2005). Corporate Scandals: The Many Faces of Greed. Obtenido de go.gale.com (Gale Academic OneFile): https://go.gale.com/ps/i.do?p=AONE&u=anon%7Eb6c97118&id=GALE|A129813945&v=2.1&it=r&sid=googleScholar&asid=2ea8b338Mart%EDnez

Rozen, C. (2014). Fraude corporativo en aumento. Revista Management 2: 1-2.