Métodos de *Machine Learning* para detección de fraude en transacciones financieras en tiempo real

Elaborado por:

Maria Camila Tarazona Nieto

Laura Carolina Mateus Agudelo

Leyla Rocío Becerra Barajas

Daniel Antonio Pérez Beltrán

Universidad EAN

Especialización en *Machine Learning*

Seminario de Investigación de Postgrado

Bogotá D.C.

Noviembre 12 de 2024

Tabla de Contenido

Métodos de *Machine Learning* para detección de fraude en transacciones financieras en tiempo real

Resumen

Las entidades financieras enfrentan un creciente desafío en la detección de fraudes en transacciones digitales debido a la sofisticación de las técnicas fraudulentas. Este estudio analiza los modelos de *Machine Learning* aplicados a la detección de fraudes en tiempo real, con el objetivo de identificar sus características, deficiencias y retos. Basado en fuentes como (Asobancaria, 2022) y la (Superintendencia Financiera, 2023), se evalúan las implicaciones prácticas y teóricas para mejorar la seguridad financiera en Colombia.

Palabras clave: *Fraude financiero, Machine Learning, detección de fraudes, entidades financieras, seguridad digital.*

Planteamiento del Problema

Las entidades financieras y sus clientes se han visto afectados por numerosos casos de fraude, cuyas técnicas de ataque son cada día más sofisticadas. Como mecanismos de prevención y detección, se han implementado varias aplicaciones de *Machine Learning*, las cuales deben ser analizadas y caracterizadas según su desempeño. Además, es importante identificar los retos para su implementación.

Antecedentes del problema.

Los casos de fraude financiero han aumentado debido al creciente uso de plataformas digitales que cada vez son más populares en Colombia. Sin embargo, la adopción de las transacciones digitales ha dado lugar a nuevos mecanismos fraudulentos, lo cual supone un desafío para las instituciones financieras, según el informe presentado por en (Superintendencia Financiera, 2023). Estas amenazas son demasiado complejas como para ser detenidas por los sistemas tradicionales de detección, que se basan en reglas predefinidas (Asobancaria, 2022). En cambio, la implementación de tecnologías avanzadas como el *Machine Learning* permite a las entidades financieras mejorar su capacidad para identificar

patrones de fraude en tiempo real. Esto ayuda a estas instituciones a adaptarse rápidamente y disminuir significativamente los índices de fraude. A pesar de que a nivel mundial empresas como el *Santander Group* han logrado una reducción significativa del fraude mediante la utilización de *Machine Learning* (Santander, 2020), en Colombia todavía existe una escasa implementación de estas tecnologías. Las instituciones financieras colombianas deben considerar optimizar sus sistemas de detección de fraude en tiempo real mediante aprendizaje automático para proteger a sus clientes y mantener la confianza en el sistema financiero dentro del entorno digital que siempre cambia (PwC, 2021).

Descripción del problema.

Una de las principales preocupaciones de las entidades financieras a nivel mundial es la detección del fraude a través de transacciones digitales. En el año 2023, el fraude en transacciones financieras representó pérdidas totales de 321.000 millones de dólares estadounidenses, con un incremento de 133% en los últimos 30 años *(Charizanos et al., 2024)*. El reto más importante en la prevención y detección del fraude en transacciones financieras es la naturaleza cambiante de los patrones y hábitos de compra de los consumidores, así como de las técnicas fraudulentas usadas por los estafadores para evitar ser detectados *(Bansal et al., 2024).* En consecuencia, se requieren mecanismos que le permitan a las entidades financieras, la prevención y detección en tiempo real de la ejecución de transacciones ilegítimas, las cuales requieren modelos que se adapten rápidamente a las estrategias de fraude cada vez más sofisticadas. Una solución que se ha venido implementando, es el uso de modelos de *Machine Learning*, los cuales pueden ser empleados para este tipo de aplicaciones, aprovechando su capacidad para aprender de datos históricos y la posibilidad de clasificación entre transacciones legítimas y fraudulentas *(Oluwabusayo et al.,2024).* Por lo anterior, es importante identificar y analizar los principales modelos que se han venido utilizando para este tipo de aplicaciones con el fin de establecer sus principales características, niveles de desempeño y retos a resolver en los próximos 5 años.

Pregunta de investigación.

¿Cómo se caracterizan los principales modelos de *Machine Learning* aplicados a la detección de fraude en transacciones financieras en tiempo real?

Objetivos

Objetivo general.

Analizar los métodos actuales de detección de fraude en tiempo real mediante *Machine Learning* con el fin de identificar sus principales características, deficiencias y retos.

Objetivos específicos.

- Realizar una búsqueda sistemática de literatura que permita determinar el estado del arte de los métodos de *Machine Learning* usados para detección de fraude en tiempo real.
- Comparar los enfoques basados en *Machine Learning* con los métodos tradicionales de detección de fraudes, analizándolos en términos de precisión y velocidad de respuesta.
- Analizar experiencias y estudios de caso de la implementación de sistemas de *Machine Learning* en otros países, extrayendo lecciones y mejores prácticas que puedan ser adaptadas al contexto de Colombia.

Justificación

El uso de *Machine Learning* (ML) como tecnología base para la detección de fraude en línea se fundamenta en su capacidad para mejorar la precisión, adaptabilidad y eficiencia de los sistemas antifraude actuales. Sin mecanismos dinámicos y evolutivos permanentes, estos sistemas corren el riesgo de quedarse rezagados ante las tácticas en constante cambio de los estafadores. Es esencial aprovechar la efectividad de los algoritmos de ML, que se entrenan a partir de características relevantes extraídas de datos históricos etiquetados, para identificar transacciones fraudulentas, incluso frente a nuevas estrategias maliciosas. Además, la implementación de ML permite automatizar la detección de fraudes, reduciendo significativamente el tiempo y los recursos necesarios para aplicar estos mecanismos en línea.

Para el sector financiero colombiano, es de gran relevancia, en especial teniendo en cuenta que en la actualidad existen más de 300 empresas que ofrecen soluciones innovadoras para

el ecosistema Fintech (Asmar, 2023). La integración de ML en estas plataformas fomenta un entorno competitivo en donde las instituciones del pueden ofrecer servicios más competitivos, ágiles y seguros.

Este proyecto se fundamenta en la necesidad crítica de comprender y mejorar las técnicas existentes frente a la creciente sofisticación de los fraudes en transacciones digitales (Superintendencia Financiera, 2023). La investigación tiene implicaciones prácticas significativas, ya que permitirá identificar las deficiencias y retos de los métodos actuales, proporcionando una base sólida para futuras mejoras en los sistemas de detección de fraude. Metodológicamente, el proyecto contribuye a la sistematización del conocimiento sobre técnicas de *Machine Learning* aplicadas a este campo, facilitando comparaciones rigurosas entre modelos avanzados y métodos tradicionales (Asobancaria, 2022).

El valor teórico del estudio radica en la identificación del estado del arte y la proyección de los desafíos futuros en la aplicación del *Machine Learning* para la detección de fraude financiero. Esto permitirá a las instituciones financieras anticipar y prepararse para los retos de los próximos años, mejorando la precisión en la detección y reduciendo los falsos positivos (PwC, 2021).

El impacto de este estudio en Colombia es significativo, considerando que, según el informe del segundo trimestre de 2024 de la Superintendencia Financiera de Colombia, se registraron 2.839 millones de transacciones por un valor total de 2.727 billones de pesos, de las cuales el 80% se realizaron a través de canales no presenciales. Entre 2023 y 2024, la cantidad de transacciones no presenciales aumentó en un 17% (Superintendencia Financiera de Colombia, 2024). En este mismo periodo, la principal queja de los consumidores financieros fue el no reconocimiento de transacciones debido a posibles fraudes, representando el 32% de las quejas presentadas ante las entidades bancarias. Por lo tanto, los resultados de este análisis serán esenciales para guiar el desarrollo e implementación de sistemas más eficientes y adaptables que protejan a los clientes y fortalezcan la confianza en el sistema financiero digital colombiano (Santander, 2020). Estos avances podrían tener efectos significativos para la banca, como la reducción de pérdidas por transacciones fraudulentas, optimización de costos operativos, fortalecimiento de la seguridad financiera y reducción de vulnerabilidades ante

ciberataques. Además, ayudarían a cumplir con normativas, mejorar la reputación de la banca y sus servicios, y fomentar nuevos ecosistemas Fintech. Para los clientes, esto significaría mayor protección de sus recursos, menos interrupciones del servicio gracias a mecanismos preventivos antifraude, respuestas rápidas ante actividades sospechosas, y una posible reducción de costos administrativos y tiempos asociados a reclamaciones, lo que fortalecería su confianza en las transacciones no presenciales.

Marco teórico

La banca se fundamenta en la realización de servicios financieros a través de transacciones, entendidas como acuerdos o movimientos entre dos o más partes donde un activo se intercambia a cambio de un pago (BBVA México, n.d.). Con el avance de la banca digital, estas transacciones pueden realizarse desde cualquier dispositivo con conexión a internet, lo que ha incrementado la exposición a riesgos de fraude digital. Este escenario ha vuelto indispensable el desarrollo de sistemas robustos de detección y prevención de fraude, especialmente para las entidades bancarias, que deben garantizar la seguridad de las transacciones en entornos digitales. Según Bello et al. (2024), el fraude financiero abarca una serie de actividades ilegales o engañosas destinadas a obtener beneficios económicos indebidos, tales como el robo de identidad, fraudes en transacciones y el lavado de dinero. Para enfrentar estos desafíos, las entidades financieras implementan tecnologías avanzadas y prácticas de seguridad que buscan proteger tanto a los usuarios como la integridad del sistema financiero en su conjunto.

De acuerdo con (Rodrigues et al., 2022), dichas prácticas incluyen procesos de detección y prevención del fraude. Los procesos de *prevención del fraude* se encargan de identificar alguna actividad sospechosa o fraudulenta y detener la ocurrencia de un fraude. Por otra parte, (Charizanos et al., 2024) define los procesos de *detección de fraude* como aquellos que se encargan de la identificación de transacciones fraudulentas mediante el análisis de patrones irregulares en tiempo real, utilizando métodos tradicionales y técnicas avanzadas como *Machine Learning*. En la figura 1 se puede observar el flujo de transacciones en línea de comercio electrónico y se destacan las zonas en donde puede aplicarse la prevención y/o la detección del fraude. Como se evidencia, una vez emitida la autorización bancaria, la prevención ya no es posible, solo la detección.

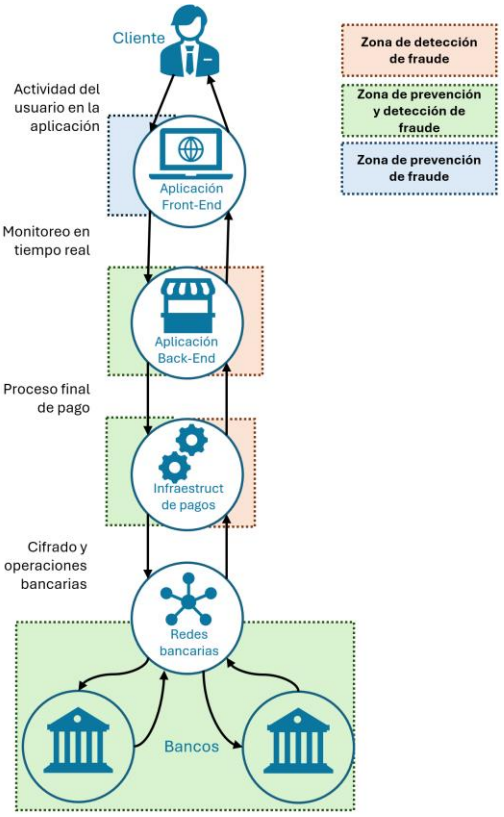

Figura 1. Zonas de detección y prevención de fraudes en un flujo de transacciones en línea. Elaboración propia basada en (Rodrigues et al., 2022)

Evolución de la Detección de Fraude en el Sector Financiero

Los objetivos de los estafadores han ido migrando hacia los principales mecanismos que soportan el comercio electrónico propiciando oportunidades para la ejecución de diferentes tipos de ataques digitales conocidos como ciberdelitos (Villamil Arcos, C, 2022). Los ciberdelitos se sustentan en diferentes métodos informáticos que permiten la captura y manipulación de los datos de las personas con el objetivo de vulnerar los sistemas financieros a través de la realización de fraude por medio de transacciones que permiten generar ganancias económicas a los atacantes, alterando el patrimonio de las víctimas. Los perpetradores de ciberdelitos son difíles de identificar y en la medida en que sus habilidades y avances tecnológicos se hacen más complejos se vuelven más difíciles de combatir (Zhang, Y. et al, 2022). Es por ello que las

Dependiendo del tipo de mecanismo a proteger, existen métodos para la detección y prevención del fraude entre ellos se puede encontrar la minería de datos, la inferencia estadística, ontologías o algoritmos de aplicación específica hasta modelos de *Machine Learning* y Deep Learning (Rodrigues et al., 2022).

El fraude en transacciones financieras ha crecido a la par del incremento en las operaciones digitales, lo que ha impulsado la evolución de las técnicas de detección. Los métodos tradicionales basados en reglas, aunque efectivos en el pasado, han quedado obsoletos frente a los sofisticados estafadores actuales.

Los sistemas basados en reglas son estáticos y dependen de parámetros predefinidos que no se actualizan automáticamente, lo que limita su capacidad para adaptarse a nuevas formas de fraude.

El fraude financiero sigue siendo un problema constante para las instituciones financieras de todo el mundo, especialmente con el aumento de las transacciones digitales. Aunque los métodos avanzados de detección han ganado popularidad, muchas organizaciones todavía dependen de enfoques tradicionales para combatir el fraude. A continuación, se describen estos métodos con sus detalles técnicos clave.

Sistemas Basados en Reglas

Los sistemas basados en reglas funcionan estableciendo condiciones específicas que determinan cuándo una transacción debería considerarse sospechosa. Estas reglas suelen incluir parámetros como:

1. Montos límite: Las transacciones que superan un valor determinado se marcan como inusuales y, potencialmente, fraudulentas. Esta regla ayuda a captar grandes sumas que podrían no ser consistentes con el comportamiento habitual del cliente.
2. Geolocalización: Si una transacción se realiza desde una región de alto riesgo o desde una ubicación distinta a la habitual del cliente, se activa una alerta. Esto es útil para evitar fraudes cuando los clientes, por ejemplo, realizan transacciones en ubicaciones remotas o inesperadas.

3. Frecuencia de transacciones: Un número elevado de transacciones en un corto período puede indicar actividad fraudulenta, especialmente si el cliente no tiene un historial de realizar múltiples operaciones en poco tiempo.

Estos sistemas son efectivos para detectar patrones básicos de fraude y suelen implementarse mediante motores de reglas que revisan cada transacción en tiempo real o en intervalos definidos. Sin embargo, una gran limitación es que los estafadores pueden aprender a evadir estas reglas ajustando sus tácticas, lo que hace necesario actualizar estos sistemas con frecuencia para mantener su efectividad (Asobancaria, 2022).

Análisis Estadístico

El análisis estadístico utiliza métodos de probabilidad y estadística para analizar patrones de transacción y detectar anomalías. Entre las técnicas estadísticas más comunes se encuentran:

1. Detección de *outliers*: Este método busca transacciones que se desvían significativamente de los patrones habituales del cliente, como un aumento abrupto en el valor de las operaciones.
2. Análisis de distribución: Emplea distribuciones estadísticas para evaluar si una transacción está dentro de los límites aceptables según el comportamiento histórico del cliente. Esto ayuda a captar operaciones inusuales.
3. Modelos de series temporales: Analiza la secuencia de transacciones a lo largo del tiempo, identificando cambios repentinos o patrones atípicos que podrían ser señales de fraude.

Para que este método funcione de manera efectiva, es necesario contar con grandes volúmenes de datos históricos y actualizar los modelos regularmente para que reflejen los patrones cambiantes de los usuarios. Sin embargo, estos modelos tienen sus limitaciones frente a fraudes complejos y suelen generar falsos positivos, ya que no siempre logran adaptarse fácilmente a nuevos comportamientos (Bansal et al., 2024; Superintendencia Financiera de Colombia, 2024).

Auditoría y Monitoreo Manual

El monitoreo manual implica que analistas especializados revisen transacciones directamente, buscando actividades sospechosas según criterios preestablecidos. Este proceso puede incluir varios enfoques:

1. Análisis contextual: Evalúa la transacción en función de detalles específicos del cliente, como su historial de transacciones o su perfil de riesgo, para decidir si es coherente con su comportamiento habitual.
2. Investigación de patrones complejos: Permite revisar múltiples transacciones que, de forma aislada, podrían parecer normales, pero en conjunto podrían indicar un fraude. Esto es especialmente útil en casos de lavado de dinero o fraude en cadena.
3. Evaluación de riesgos en transacciones de alto valor: Cuando se trata de transacciones grandes, los analistas aplican criterios adicionales de revisión y pueden solicitar documentación o verificaciones adicionales.

Aunque la revisión manual permite un análisis detallado y adaptado al contexto de cada transacción, su principal desventaja es el costo en tiempo y recursos. Escalar este método en instituciones con altos volúmenes de transacciones resulta complicado y costoso (Villamil Arcos, 2022; Bello & Komolafe, 2024).

Autenticación y Verificación de Identidad

Para evitar accesos no autorizados, las instituciones financieras están adoptando métodos avanzados de autenticación y verificación de identidad. Entre los más comunes están:

1. Autenticación en dos pasos (2FA): Combina algo que el usuario sabe (una contraseña) con algo que tiene (un código de verificación) o algo que es (biometría). Esto proporciona una capa adicional de seguridad que dificulta el acceso a personas no autorizadas.
2. Biometría: Utiliza características físicas únicas del usuario, como huellas dactilares, reconocimiento facial o escaneo de iris. Este enfoque es muy seguro, pero requiere infraestructura avanzada y puede ser costoso.

3. Autenticación basada en comportamiento: Monitorea el comportamiento del usuario en tiempo real, como la velocidad de escritura o los patrones de clics, y evalúa si coincide con el comportamiento habitual del usuario.

Estos métodos reducen considerablemente el riesgo de accesos no autorizados, aunque la implementación varía dependiendo de la capacidad tecnológica de cada institución. Las contraseñas por sí solas se están volviendo insuficientes, debido a que son vulnerables a ataques de fuerza bruta y de ingeniería social (Santander, 2020; PwC, 2021).

Listas Negras (*Blacklisting*)

Las listas negras son una herramienta efectiva que permite bloquear cuentas, tarjetas o direcciones IP asociadas a actividades fraudulentas en el pasado. Este sistema funciona de la siguiente manera:

1. Identificación de patrones de fraude conocidos: Las cuentas o IPs involucradas en fraudes previos se bloquean de inmediato en futuras transacciones, reduciendo el riesgo de reincidencia.
2. Actualización en tiempo real: Las listas negras se actualizan constantemente con datos de nuevas amenazas, permitiendo una reacción rápida ante nuevos riesgos.
3. Aplicación en múltiples puntos de control: Estas listas pueden bloquear transacciones en diferentes momentos del proceso, incluso antes de autorizar una operación, mejorando la protección contra fraudes.

Si bien las listas negras son útiles para detener amenazas conocidas, tienen sus limitaciones, especialmente cuando los defraudadores usan cuentas legítimas comprometidas o emplean técnicas de ingeniería social para eludir la detección. En estos casos, la lista negra puede no detectar el fraude (Deloitte, 2022; KPMG, 2023).

Modelos Avanzados de *Machine Learning*

Machine Learning ha surgido como una solución clave, permitiendo que las máquinas aprendan de datos históricos y se adapten a nuevas formas de fraude casi en tiempo real (Charizanos et al, 2024). El cambio hacia un enfoque basado en datos ha permitido a las

instituciones financieras abordar con mayor precisión el fraude, que históricamente se limitaba a sistemas reactivos, en lugar de proactivos. Mientras que los métodos tradicionales detectan fraudes únicamente después de que ocurren, los modelos de *Machine Learning* permiten una detección predictiva, identificando patrones inusuales en tiempo real. Esto ha sido posible gracias al aumento de la capacidad de procesamiento de datos y al acceso a grandes volúmenes de información, que alimentan estos sistemas inteligentes (Ghosh et al, 2020).

Machine Learning es una rama de la inteligencia artificial que permite a las máquinas aprender de datos y mejorar sus predicciones sin intervención humana directa (Goodfellow et al., 2016). Estos sistemas dinámicos permiten una detección más eficiente, adaptándose continuamente a los cambios en las tácticas de los defraudadores, que evolucionan rápidamente gracias al acceso a tecnologías avanzadas (Ngai et al., 2019).

Los algoritmos de *Machine Learning* están basados en dos tipos de aprendizaje: supervisado y no supervisado. Cada uno de ellos ofrece distintas técnicas como se ve en la Figura 2.

Modelos como las Máquinas de Soporte Vectorial (SVM), los Árboles de Decisión y técnicas más avanzadas como *Random Forest* han demostrado ser herramientas poderosas para identificar transacciones sospechosas. A diferencia de los métodos tradicionales basados en reglas, que se limitan a detectar fraudes conocidos mediante parámetros predefinidos, estos algoritmos de *Machine Learning* permiten detectar patrones ocultos en los datos que no son evidentes para los humanos, pero que los sistemas pueden identificar con facilidad (Bansal et al., 2024). Además, cuando no se dispone de datos etiquetados, los modelos no supervisados, que detectan patrones atípicos, son especialmente útiles. Esto es crucial en la identificación de fraudes emergentes, donde no hay ejemplos previos claros para entrenar un modelo supervisado.

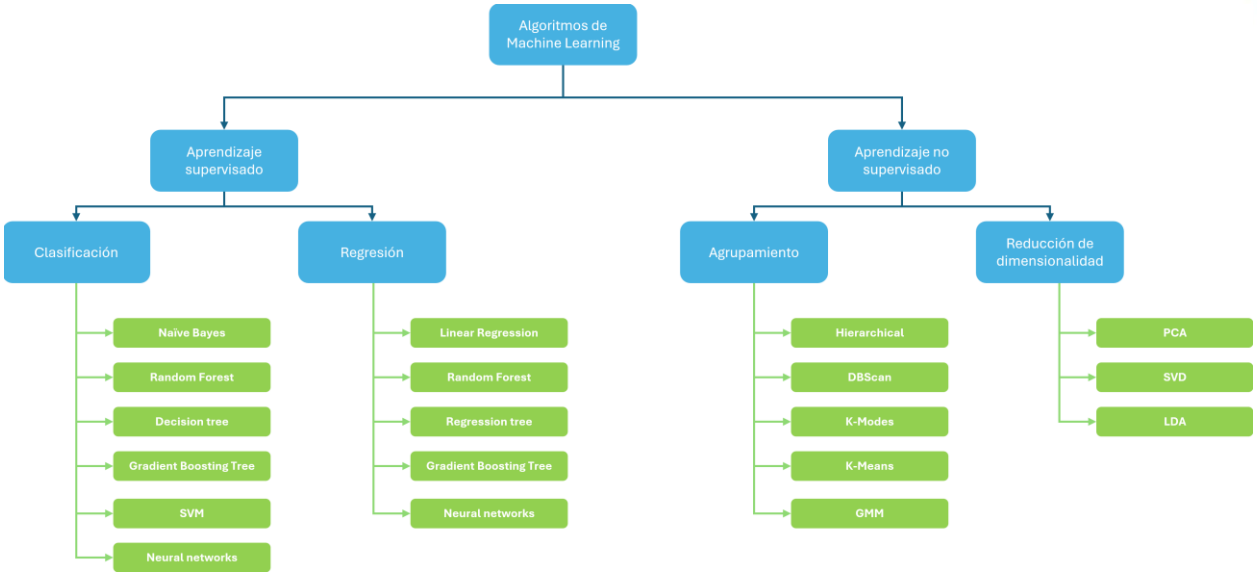

Figura 2. Tipos de algoritmos de *Machine Learning*.
Elaboración propia basada en (Villamil Arcos, C, 2022)

Un área de desarrollo interesante ha sido la combinación de diferentes enfoques a través del aprendizaje híbrido. El uso de enfoques como los sistemas de votación o ensamblado, que combinan las salidas de múltiples modelos, ha demostrado aumentar la precisión en la detección de fraudes, reduciendo los falsos positivos, un problema recurrente en los métodos tradicionales. Esto no solo mejora la detección, sino que también reduce el impacto negativo sobre los usuarios legítimos, un aspecto clave en la experiencia del cliente (Phua et al., 2010).

Funcionamiento del Machine Learning en la Detección de Fraude

El *Machine Learning* se basa en el análisis de grandes volúmenes de datos para tomar decisiones inteligentes sin la necesidad de una programación explícita para cada escenario. En la detección de fraude financiero, esto implica que los modelos pueden identificar patrones ocultos que sugieren actividades sospechosas. Por ejemplo, los Árboles de Decisión son fáciles de interpretar y proporcionan reglas claras que describen cómo se detecta el fraude, mientras que las Redes Neuronales, aunque más complejas, son capaces de detectar relaciones más sutiles en los datos (Oluwabusayo et al, 2024). Estas relaciones pueden incluir secuencias de tiempo, combinaciones de montos de transacción o ubicaciones geográficas inusuales, que son señales claras de posible fraude.

Aunque las Redes Neuronales han mostrado un gran potencial en la detección de fraudes, requieren grandes volúmenes de datos y un alto poder de procesamiento. Un caso particular de interés son las Redes Neuronales Recurrentes (RNN), que son efectivas en el análisis de series temporales, como las transacciones financieras, ya que tienen la capacidad de "recordar" patrones previos en la secuencia de datos (Goodfellow et al, 2016). Esta capacidad es particularmente útil para detectar cambios en el comportamiento del usuario a lo largo del tiempo, lo que permite una identificación más precisa de actividades anómalas.

Por otro lado, los modelos no supervisados, como las Redes Bayesianas, son ideales para identificar transacciones fuera de lo común en entornos con poca información previa (McKinsey et al, 2021). Estos modelos han permitido la detección de fraudes completamente nuevos, que no habrían sido detectados por sistemas basados en reglas predefinidas, mejorando así la capacidad de respuesta de las instituciones financieras ante amenazas inesperadas.

Modelos de Machine Learning para la detección de fraude en tiempo real

Para diseñar una solución eficaz de detección de fraude en transacciones financieras en tiempo real, es ideal emplear un enfoque híbrido que combine tanto modelos supervisados como no supervisados. Este enfoque tiene la ventaja de identificar patrones de fraude conocidos, mientras detecta también patrones nuevos y emergentes.

Modelos Supervisados

1. *Random Forest*: Este modelo clasificador se basa en una colección de árboles de decisión, donde cada uno clasifica de forma independiente. Los árboles se construyen utilizando diferentes subconjuntos de características y datos, lo que permite detectar patrones de fraude que están bien definidos, como montos altos y ubicaciones atípicas. Al promediar los resultados de varios árboles, *Random Forest* mejora tanto la precisión como la estabilidad, minimizando el sobreajuste y permitiendo generalizar mejor los datos de entrada.
2. Redes Neuronales Profundas (Deep Neural Networks): Estas redes son poderosas para captar patrones complejos en los datos y son especialmente útiles en escenarios donde

las transacciones fraudulentas no son fácilmente distinguibles. La estructura multicapa de las redes neuronales profundas permite encontrar relaciones no lineales entre las variables. Si bien estas redes requieren más datos y poder computacional para su entrenamiento, ofrecen una precisión sobresaliente en la detección de fraudes sofisticados.

Modelos No Supervisados

1. *Isolation Forest*: Este modelo detecta anomalías al analizar la "facilidad" con la que una transacción puede ser aislada del conjunto de datos. Las transacciones que presentan patrones inusuales se aíslan rápidamente, señalando posibles fraudes. Este modelo es eficiente y adecuado para conjuntos de datos grandes, ya que identifica comportamientos sospechosos sin necesidad de etiquetar los datos.
2. *Autoencoders*: Los *autoencoders* son redes neuronales que intentan reconstruir sus entradas a partir de capas ocultas. Al analizar el error de reconstrucción, pueden detectar transacciones que se desvían significativamente de los patrones esperados. Si una transacción muestra un error de reconstrucción alto, podría tratarse de una anomalía. Este enfoque es particularmente útil cuando se busca detectar fraudes con patrones complejos y poco evidentes.

Modelos de Aprendizaje Secuencial (para series temporales)

1. Redes LSTM (*Long Short-Term Memory*): Las LSTM son un tipo especializado de red neuronal recurrente que "recuerda" información pasada a lo largo de una secuencia de datos, permitiendo analizar patrones de tiempo en las transacciones. Esto resulta muy útil para detectar cambios repentinos en el comportamiento de los usuarios, como un aumento en la frecuencia de transacciones o un cambio en las ubicaciones de uso. Las LSTM son ideales para detectar anomalías en patrones temporales, como transacciones fuera de horario habitual.

La razón para elegir un enfoque híbrido es su flexibilidad y adaptabilidad, que lo hacen adecuado tanto para fraudes tradicionales como emergentes. Mientras que los modelos supervisados se enfocan en patrones históricos, los no supervisados permiten detectar

anomalías en transacciones nuevas. Las LSTM, al trabajar con series temporales, añaden una capa crítica de análisis en tiempo real. Esta combinación no solo optimiza la precisión, sino que también reduce los falsos positivos, un factor clave en la experiencia del cliente y en la credibilidad del sistema de detección.

Elementos Fundamentales para la Calidad y Rendimiento del modelo

La aplicación de *Machine Learning* (ML) al análisis de datos implica entrenar modelos con conjuntos de datos (*datasets*) cuidadosamente seleccionados y preparados para asegurar resultados precisos y efectivos. La calidad y las características de un *conjunto de datos* juegan un papel crucial en el rendimiento del modelo, ya que los algoritmos de ML aprenden patrones y relaciones a partir de la información que se les proporciona. Un *conjunto de datos* ideal debe tener ciertas características clave. Primero, es fundamental que los datos sean de calidad y estén limpios, es decir, libres de errores, inconsistencias y valores atípicos extremos. Esto minimiza el ruido y permite que los modelos aprendan de manera más eficiente. Además, la variedad y representatividad del *conjunto de datos* es esencial para reflejar adecuadamente el problema que se desea resolver, lo cual incluye una diversidad de casos y escenarios que evitan sesgos y aseguran una buena generalización del modelo.

Otra característica importante es el volumen suficiente de información. Dependiendo de la complejidad del modelo y la naturaleza del problema, se necesita un volumen considerable para que el modelo pueda capturar patrones con confianza. En problemas de clasificación, como la detección de fraudes, el equilibrio de clases es crucial. Un *conjunto de datos* desbalanceado, donde una clase esté sobrerrepresentada, puede llevar a que el modelo prediga solo la clase mayoritaria y falle en identificar la clase minoritaria. Además, la relevancia de las características o variables incluidas en el *conjunto de datos* es vital. La ingeniería de características y la selección de las más significativas mejoran la precisión y la interpretabilidad del modelo.

La importancia de un buen *conjunto de datos* no puede subestimarse. Un *conjunto de datos* bien diseñado y estructurado es fundamental para que el modelo de ML entrene de forma adecuada y generalice sus predicciones en datos nuevos y desconocidos. La calidad del *conjunto de datos* influye directamente en el rendimiento y la capacidad del modelo para extraer

patrones precisos y tomar decisiones confiables. En resumen, el éxito de un modelo de *Machine Learning* depende en gran medida de la preparación y la calidad del *conjunto de datos*. Un enfoque cuidadoso en la selección, limpieza y balance de los datos garantiza que los modelos produzcan resultados útiles y precisos en el contexto en el que se aplican.

Metodología

Primer nivel

Este trabajo se desarrollará a partir de un enfoque metodológico cualitativo, centrado en una revisión sistemática de la literatura. El objetivo es explorar y sintetizar el estado del arte en la aplicación de *Machine Learning* para la detección de fraudes financieros en tiempo real, identificar los principales tipos de fraude y caracterizar las técnicas y enfoques más utilizados para su detección en línea. De este modo, se consolidará un estudio descriptivo que ofrecerá una visión integral del tema y que, además, se alineará con casos de uso potenciales en el contexto colombiano.

Segundo nivel

La metodología está basada en varias etapas, de acuerdo con lo mostrado en la Figura 3. Se contemplan tres etapas:

1. Revisión sistemática de literatura
2. Identificación de principales modelos aplicados a la detección en línea de fraude en transacciones financieras
3. Alineación al caso colombiano

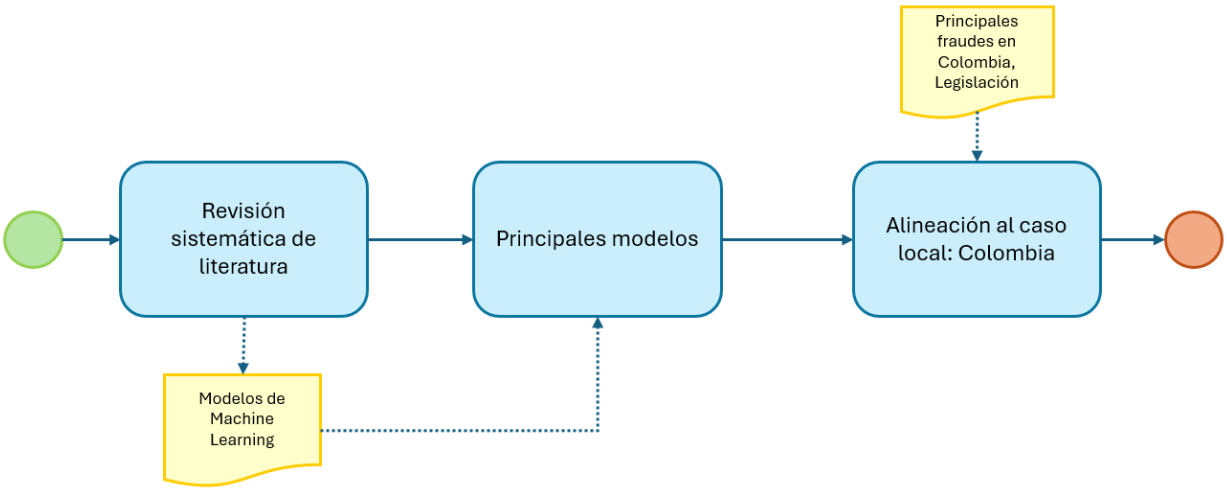

Figura 3: Proceso metodológico. Elaboración propia

1. Revisión sistemática de literatura:

Los instrumentos definidos para la recolección de la información están basados en la revisión sistemática de literatura (Kitchenham et al., 2008), la cual sigue una serie de pasos:

1.1. Planificación:

La etapa de planificación permite identificar los temas y conceptos que serán usados para la búsqueda de literatura los cuales son expresados a través de *"preguntas orientadoras"*. Para este caso se usarán:

a) *¿Cuáles son los principales modelos y algoritmos de Machine Learning utilizados en la detección de fraudes en línea en el sector financiero a nivel global?*
b) *¿Qué técnicas de Machine Learning han demostrado ser más efectivas para detectar fraudes en tiempo real?*
c) *¿Cómo se comparan los enfoques basados en Machine Learning con los métodos tradicionales de detección de fraudes en términos de precisión y velocidad?*
d) *¿Qué lecciones se pueden extraer de la implementación de sistemas de Machine Learning en la detección de fraudes en otros países que puedan ser aplicables a Colombia?*

1.2. Búsqueda

Esta etapa implica la identificación y recopilación de estudios relevantes a partir de estudios científicos relevantes. Para ello se tendrán en cuenta dos aspectos:

Fuentes de información:

Como fuentes principales de información y de obtención de documentos relevantes se usaron *Elsevier Scopus*.

Ecuación de búsqueda

Con el fin de realizar las consultas en la fuente seleccionada se elabora una estrategia de búsqueda mediante una ecuación, la cual está conformada por las palabras que se consideran claves para reducir la búsqueda a los documentos más significativos. Así mismo, se establece mediante ella los criterios de inclusión y exclusión precisos. A continuación, la ecuación de búsqueda seleccionada:

(TITLE-ABS-KEY (online AND fraud AND detection AND "machine learning") AND LANGUAGE (English)) AND PUBYEAR > 2020

Para la alineación de los casos de uso en Colombia, se utiliza la ecuación de búsqueda:

(TITLE-ABS-KEY (fraude AND transacciones AND financieras AND Colombia) AND LANGUAGE (English OR Spanish)) AND PUBYEAR > 2020

Finalmente, en esta etapa se realiza la búsqueda automática de los estudios científicos para la revisión cualitativa.

1.3. Selección rigurosa de los estudios

Mediante esta etapa se busca llevar a cabo una *lectura inicial o de primer nivel* (lectura del título, resumen y principales capítulos de cada artículo) y selección de los estudios considerando únicamente aquellos que describen o responde a alguna de las preguntas

orientadoras. La organización bibliográfica que facilitará el análisis avanzado de los documentos se soportará mediante la herramienta ***Mendeley***.

1.4. Evaluación de calidad

Se procede a realizar una evaluación de calidad de los estudios que fueron seleccionados en la etapa anterior, según su rigor metodológico. Esto implica un análisis crítico de cómo se llevó a cabo cada estudio, la validez de sus conclusiones y la relevancia para la pregunta de investigación.

1.5. Interpretación e integración

Finalmente, se procede a resumir los hallazgos que se pueden extraer de los estudios seleccionados, realizando un análisis identificar patrones comunes que sustenten la pregunta de investigación de este trabajo, identifiquen aspectos emergentes que no se hayan identificado previamente y que permita evidenciar oportunidades para trabajos futuros.

2. Principales modelos

Como resultado de la etapa anterior, se espera poder extraer información descriptiva de los modelos de *Machine Learning* que pueden ser aplicados a la detección de fraude financiero en línea y los principales aspectos que permiten su caracterización. Esta información es la entrada de esta etapa del proceso, en donde el objetivo será tabular los modelos identificados describiendo: características clave, aplicaciones y referencias.

3. Alineación al caso colombiano

Con el fin de identificar qué modelos utilizados en estudios globales podrían aplicarse al contexto local, se analizarán los principales tipos de fraude que se presentan en Colombia, así como los desafíos tecnológicos y regulatorios específicos de la detección de fraude en transacciones financiera. A partir de la caracterización previa, se seleccionarán los métodos de *Machine Learning* más adecuados para abordar estos retos en el entorno colombiano.

Búsqueda sistemática de literatura

Análisis bibliográfico

Mediante la metodología propuesta de búsqueda sistemática de literatura, y utilizando como fuente principal de datos Elsevier Scopus, se realizó una primera búsqueda de información usando la ecuación de búsqueda propuesta ver *Ecuación 1.* El resultado fueron 660 artículos publicados desde el año 2020 hasta la actualidad. De acuerdo con la figura 4, se evidencia el creciente interés de la comunidad académica en el tema.

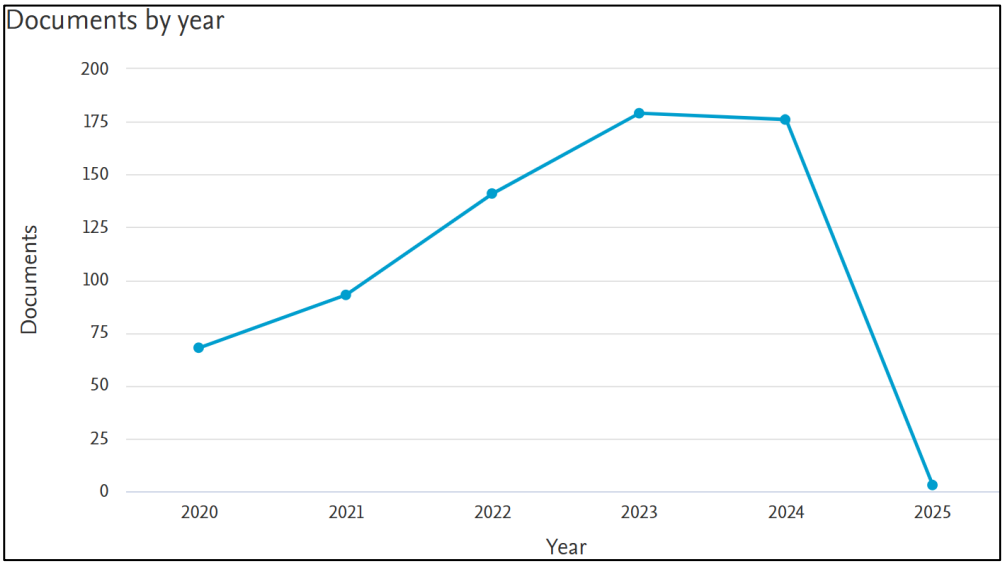

Figura 4. Documentos publicados por año Fuente *Elsevier SCOPUS*

En la Figura 5, es posible evidenciar que el país en donde se ha producido casi el 50% de documentos científicos sobre la detección de fraude en línea mediante el uso de *Machine Learning* es India, seguido por China, Estados Unidos y Arabia Saudita.

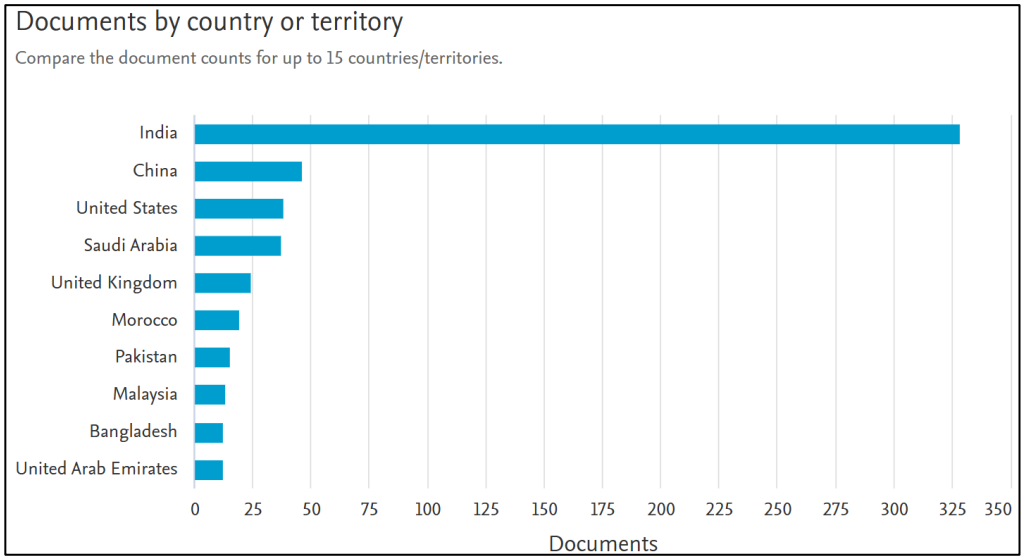

Figura 5. Documentos publicados por país. Fuente *Elsevier SCOPUS*

Un análisis bibliográfico realizado mediante la herramienta *VOSViewer,* permitió la creación de algunos mapas de interés para encontrar patrones entre los documentos analizados. Ellos son:

1. ***Mapa de red de términos de autor***: desplegado en la Figura 6, sirve para identificar la forma como se relacionan los principales términos clave en la investigación. En este caso los términos más relevantes son "*Machine Learning*", "fraud detection", "*Random Forest*", "*credit card*", "*classification*". La conexión entre ellos permite establecer la relación entre los conceptos y la definición de algunos clústeres o subtemas como:
    - "*Machine Learning* – Deep Learning – Classification" resaltado en verde, el cual evidencia preliminarmente el uso de redes neuronales y aprendizaje profundo en mecanismos de clasificación en detección de fraude.
    - "Fraud detection – xgboost – ecommerce – anomaly detection" resaltado en azul, que indica el enfoque en detección de fraudes y anomalías.
    - "*Random Forest* – logistic regression – decision tree – credit card" resaltado en rojo, que está asociado al uso de algoritmos específicos en fraude con tarjetas de crédito.

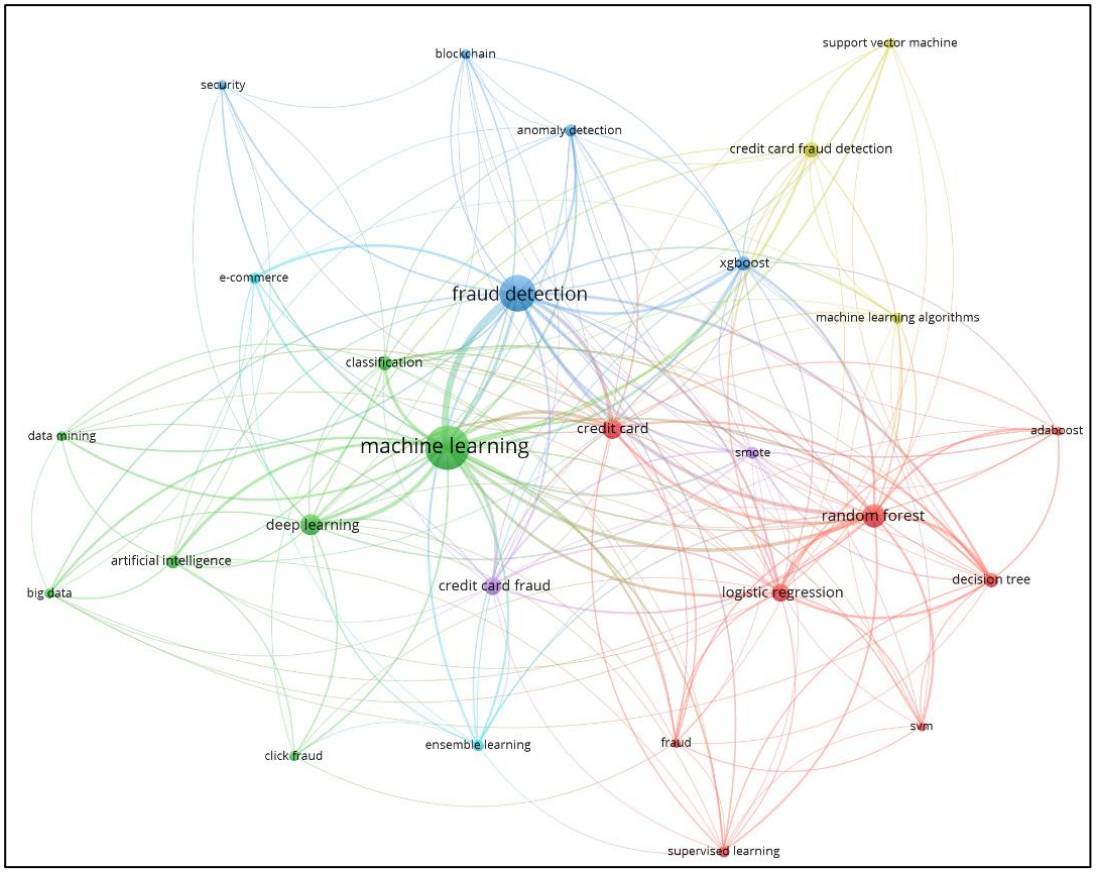

Figura 6. Mapa de red de términos usados por los autores. Fuente *VOSViewer*

2. ***Mapa de red de términos de acoplamiento bibliográfico,*** desplegado en la Figura 7, permite identificar cómo los documentos se conectan entre sí en términos de referencias comunes, su evolución temporal de acuerdo con el color del nodo, documentos más citados de acuerdo con el tamaño del nodo y la relación entre varios de ellos a través del grosor del vínculo. En este caso, se evidencia que muchos de los artículos más citados fueron escritos en el año 2020 y la gran mayoría en el año 2022. Se evidencia como fuentes más citadas Taha (2020), Dhieb (2020), Alzaharani (2021), entre otros.

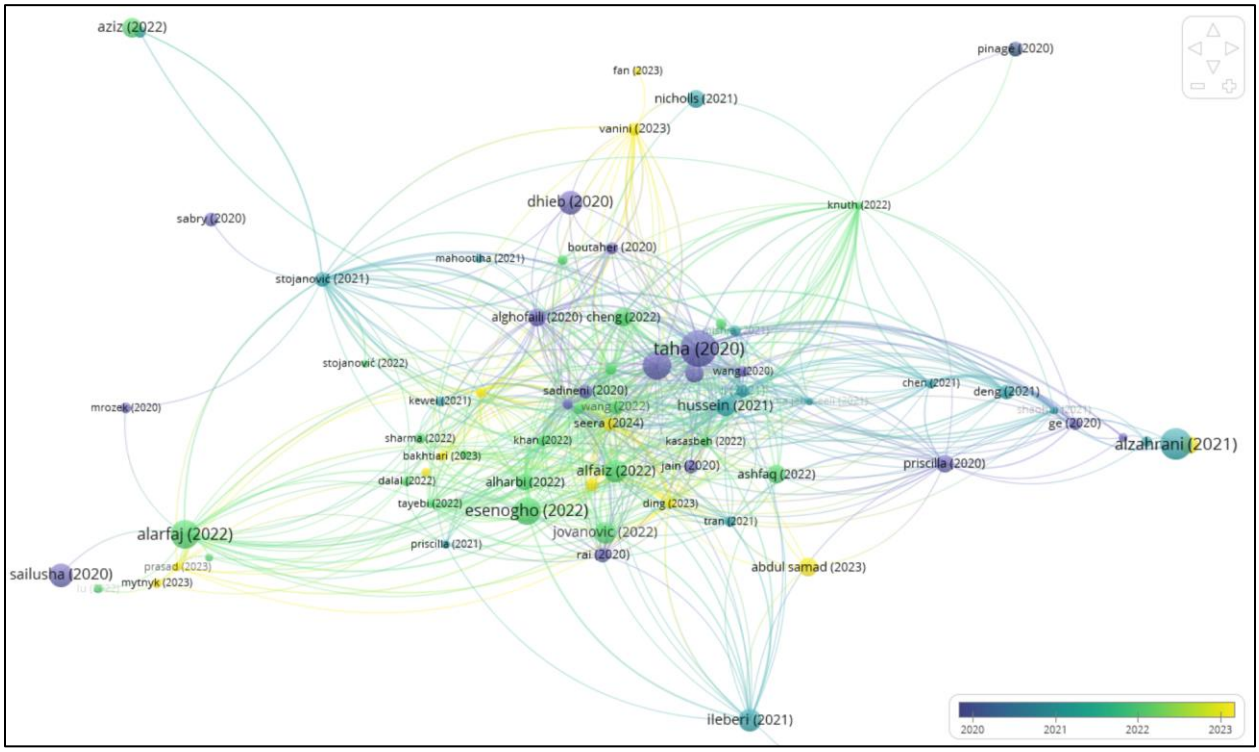

Figura 6. Mapa de red de acoplamiento bibliográfico por año. Fuente *VOSViewer*

Selección de artículos de acuerdo con criterios de inclusión/exclusión

La selección de los estudios se realizó mediante los criterios de inclusión y excusión mostrados en la Tabla 1.

Aspecto	**Criterios**	
	Inclusión	**Exclusión**
Año de publicación del estudio	>= 2020	<2020
Idioma	Inglés, español	Otros idiomas
Tipo de estudio	Artículo científico Conferencia Tesis Revista Informes técnicos	Páginas web con información no sustentada

Aspecto	Criterios	
	Inclusión	**Exclusión**
Cantidad de citaciones	Mayores a 5 citaciones para estudios de antes de 2022.	Sin citaciones
Temas tratados en el artículo	Aplicación de *Machine Learning* en transacciones financieras en línea	Artículos que no se centran en el uso de *Machine Learning* en transacciones financieras en línea

Tabla 1. Criterios de inclusión y exclusión

Los resultados obtenidos durante la etapa de revisión de la literatura y la selección de estudios relevantes permitieron identificar los artículos que servirán de base para responder a las preguntas de investigación planteadas en el marco metodológico.

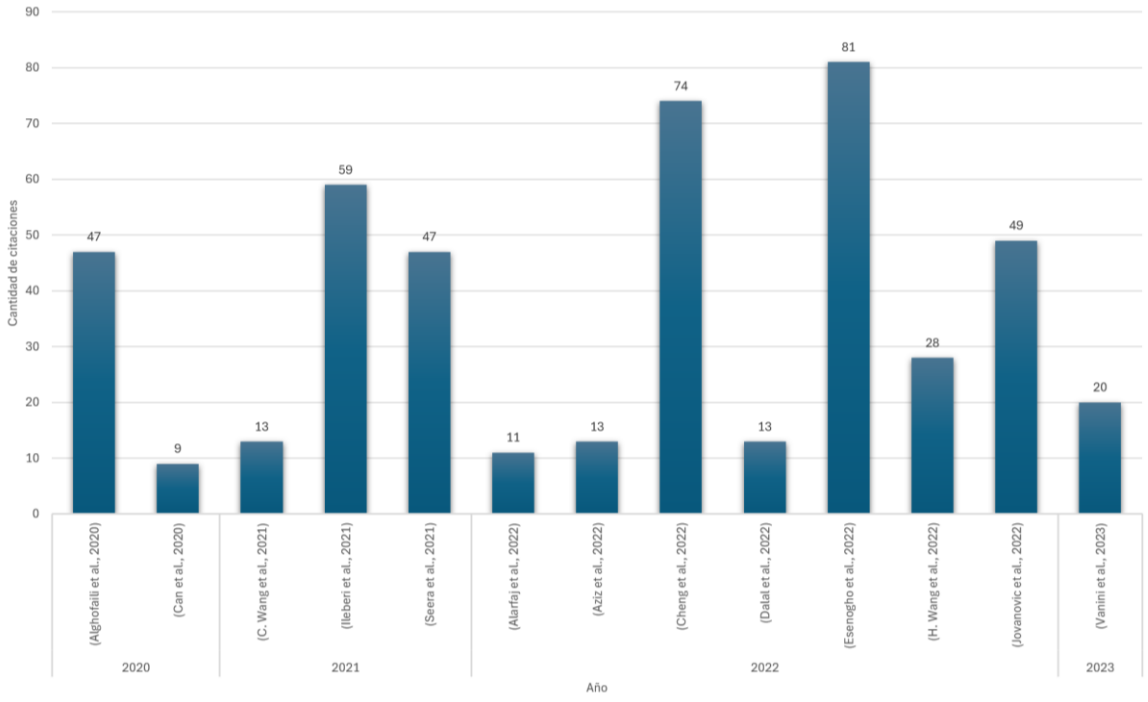

Figura 7. Estudios más relevantes

Estos estudios aportan una perspectiva detallada del estado del arte en cuanto a los modelos de *Machine Learning* utilizados para la detección de fraudes en el sector financiero. Los artículos más destacados, seleccionados en función de su relevancia y la cantidad de citas recibidas, se presentan en la Figura 7.

Estado del arte

La investigación reciente sobre el uso de aprendizaje automático para la detección de fraudes en transacciones financieras en línea ha mostrado los esfuerzos por aplicar distintos enfoques para mejorar el desempeño de los sistemas de detección de fraude tradicionales y/o aquellos con modelos de *Machine Learning* menos precisos y con menor desempeño. Algunos estudios han empleado técnicas de aprendizaje profundo, incluidas arquitecturas basadas en LSTM (Alghofaili et al., 2020) y CNN (Alarfaj et al., 2022), demostrando alta precisión en la detección de fraudes. Las redes neuronales gráficas con atención espacio-temporal han mostrado ser prometedoras para capturar patrones de fraude complejos (Cheng et al., 2022). Otros enfoques incluyen ventanas de tiempo deslizantes adaptativas (Wang et al., 2021), métodos de ensamble como AdaBoost (Ileberi et al., 2021) y Light GBM optimizado (Aziz et al., 2022). Los investigadores también han investigado el aprendizaje automático cuántico, que superó a los métodos tradicionales en datos de series temporales altamente desbalanceados (Wang et al., 2022). Para abordar el desbalance de clases, se han aplicado técnicas como SMOTE (Ileberi et al., 2021). Más allá de la detección, algunos estudios se han centrado en la gestión de riesgos y la optimización económica de las medidas de prevención de fraudes (Vanini et al., 2023). Estos avances buscan mejorar la precisión, la velocidad y la adaptabilidad de la detección de fraudes en escenarios del mundo real.

Pregunta 1: ¿Cuáles son los principales modelos y algoritmos de Machine Learning utilizados en la detección de fraudes en línea en el sector financiero a nivel global?

La detección de fraudes financieros en línea ha recibido una atención creciente por parte de diversos autores a nivel global, como se muestra en la Tabla 2. Esta tabla resume la técnica

empleada en cada estudio, el tipo de fraude abordado y el *conjunto de datos* utilizado para entrenar los modelos. Esta información proporciona una visión comparativa y facilita la identificación de enfoques que podrían adaptarse a contextos específicos.

Autor	Técnica empleada	Conjunto de datos
(Alghofaili et al., 2020)	Implementa un modelo de aprendizaje profundo mediante *LSTM* cuyos resultados fueron comparados con un modelo existente de *Autoencoder*.	*Conjunto de datos* real de fraudes con tarjetas de crédito
(Can et al., 2020)	El artículo examina las características de las transacciones fraudulentas con tarjetas para desarrollar modelos de detección de fraude basados en aprendizaje automático más resilientes y se usó para desarrollar y evaluar diferentes modelos de detección de fraude basados en perfiles.	El conjunto de datos utilizado en este estudio consta de 4.000 millones de transacciones no fraudulentas y 245.000 transacciones fraudulentas, lo que supone un total de más de 4.200 millones de transacciones. El conjunto de datos fue aportado por 35 bancos de Turquía.
(C. Wang et al., 2021)	Usa un enfoque de aprendizaje adaptativo que adapta una ventana automática de aprendizaje (LAW) para ajustar los parámetros de la ventana de tiempo para la detección de fraude en línea.	*Conjunto de datos* real del servicio de pago en línea de un banco comercial.
(Ileberi et al., 2021)	Usa una técnica de *Boosting Adaptativo* (*AdaBoost*) combinada con algoritmos de aprendizaje automático para la detección de fraudes con tarjetas de crédito.	Usa dos conjuntos de datos: 1) Un conjunto de datos real, desequilibrado, de transacciones con tarjetas de crédito de titulares europeos.

Autor	Técnica empleada	Conjunto de datos
		2) Un conjunto de datos sintético y altamente sesgado de fraudes con tarjetas de crédito.
(Seera et al., 2021)	Usa modelos estadísticos y de aprendizaje automático para detección de fraude con tarjetas de pago.	Usa un *conjunto de datos* de registros reales de acceso público
(H. Wang et al., 2022)	Usa aprendizaje automático mejorado con computación cuántica lo cual mejorar sustancialmente la velocidad y la precisión de la detección de fraude en línea, especialmente para datos de series de tiempo altamente desequilibrados.	Los *conjuntos de datos* son: 1. Transacciones con tarjetas de crédito de Israel (serie no temporal) que presentan un desequilibrio moderado 2. Un conjunto de datos de préstamos bancarios (serie temporal) que presenta un desequilibrio Elevado
(Alarfaj et al., 2022)	Propone varias técnicas: - Algoritmos de aprendizaje automático: método de aprendizaje extremo, árbol de decisiones, bosque aleatorio, máquina de vectores de soporte, regresión logística, XG Boost - Algoritmos de aprendizaje profundo: arquitecturas de redes neuronales convolucionales con diferentes cantidades de capas ocultas	El conjunto de datos utilizado consiste en información pública de transacciones de dos días en Septiembre de 2018 usado para detección de fraudes con tarjetas de crédito. El conjunto de datos tiene características de alto desequilibrio de clases, patrones de fraude cambiantes y altas tasas de falsas alarmas, lo que lo

Autor	Técnica empleada	Conjunto de datos
		convierte en un conjunto de datos del mundo real desafiante para la detección de fraudes con tarjetas de crédito.
(Cheng et al., 2022)	Usa una red neuronal gráfica basada en la atención espacio-temporal y redes neuronales convolucionales para la detección de fraudes con tarjetas de crédito con mejoras sustanciales respecto a otros métodos de última generación.	Conjunto de datos de transacciones con tarjetas del mundo real que se utilizó para entrenar y evaluar el modelo STAGN propuesto para la detección de fraudes con tarjetas de crédito.
(Aziz et al., 2022)	Un enfoque de aprendizaje automático para detectar eficazmente las transacciones fraudulentas de Ethereum. Las técnicas propuestas son: - *Light Gradient Boosting Machine* (LGBM) con optimización de la distancia euclidiana - Ajuste de hiperparámetros mediante estimación estructurada de la distancia euclidiana - Comparación de LGBM con otros modelos de aprendizaje automático como Random Forest, Logistic Regression, MLP, KNN, XGBoost, SVC y AdaBoost	El conjunto de datos incluye 9841 transacciones o filas de Ethereum que se han identificado como fraudulentas o legítimas en el sitio web de Kaggle.
(Jovanovic et al., 2022)	Propone un enfoque híbrido de aprendizaje automático y metaheurística de enjambre para la detección de fraudes	El conjunto de datos utilizado en el estudio es el conjunto de datos de detección de fraudes con

Autor	Técnica empleada	Conjunto de datos
	con tarjetas de crédito. Usa modelos SVM y ELM.	tarjetas de crédito, que consta de transacciones con tarjetas de crédito generadas en Europa en septiembre de 2013 durante un período de dos días. El conjunto de datos está muy desequilibrado, con solo 492 transacciones fraudulentas de un total de 284.807 transacciones, lo que representa solo el 0,172% del conjunto de datos.
(Esenogho et al., 2022)	Usa un conjunto de redes neuronales el cual es ajustado mediante ingeniería de características para mejorar la detección de fraudes con tarjetas de crédito. Aprovecha las características de AdaBoost con LSTM como mecanismo base.	Conjuntos de datos de transacciones con tarjeta de crédito reales disponibles al público.
(Dalal et al., 2022)	Implementa una técnica híbrida que combina el ajuste de hiperparámetros inspirado en la naturaleza y XGBoost para detectar el fraude en los pagos financieros.	El conjunto de datos utilizado en el estudio es el conjunto de datos Banksim, los cuales son generados sintéticamente.
(Vanini et al., 2023)	El estudio propone los modelos de aprendizaje automático como mecanismos eficaces en la detección del fraude en los pagos en línea. Valida la optimización económica de los resultados	*Conjunto de datos* real de transacciones de pago en línea, que incluye información sobre fraudes, pérdidas y tasas de falsos positivos.

Autor	Técnica empleada	Conjunto de datos
	de su aplicación y propone un modelo para predecir el riesgo de fraude.	

Tabla 2. Técnicas de *Machine Learning* más usadas recientemente a nivel global

Pregunta 2: ¿Qué técnicas de Machine Learning han demostrado ser más efectivas para detectar fraudes en tiempo real?

Las técnicas más efectivas para detectar fraudes en tiempo real incluyen:

1. Modelos basados en anomalías (Isolation Forest, One-Class SVM), que identifican transacciones inusuales sin etiquetas previas.
2. Algoritmos de clasificación supervisados (Random Forest, Gradient Boosting), útiles para aprender patrones específicos de fraude a partir de datos históricos.
3. Modelos de secuencias temporales (LSTM, GRU), que detectan cambios en los patrones de comportamiento en transacciones a lo largo del tiempo.
4. Técnicas de clustering no supervisado (K-Means), que agrupan transacciones similares, resaltando las que son anómalas.

Pregunta 3: ¿Cómo se comparan los enfoques basados en Machine Learning con los métodos tradicionales de detección de fraudes en términos de precisión y velocidad?

Precisión: Los métodos tradicionales, que usan reglas y estadística, suelen ser menos precisos, especialmente ante fraudes nuevos o más complejos (Ngai et al., 2019).

Por otro lado, los modelos de *Machine Learning*, como Bosques Aleatorios y Redes Neuronales, son más precisos porque pueden aprender patrones complejos y adaptarse a cambios en los datos (Bansal et al., 2024).

Velocidad: Los métodos tradicionales suelen ser más lentos debido a su dependencia de procesos manuales (Sharma & Panigrahi, 2022).

Algoritmos de ML, como Isolation Forest y XGBoost, procesan datos en tiempo real, detectando transacciones sospechosas de manera inmediata (Bello & Komolafe, 2024).

Adaptabilidad: Los métodos tradicionales solo funcionan bien con fraudes conocidos, mientras que el *Machine Learning* puede aprender nuevas tácticas de fraude y adaptarse, lo que los hace más efectivos (Mirkovic & Yang, 2023).

En general, los enfoques de *Machine Learning* superan a los métodos tradicionales en la detección de fraudes. Son más precisos porque pueden reconocer patrones complejos, más rápidos gracias a su capacidad de procesar datos en tiempo real, y más adaptables a las nuevas tácticas de fraude.

Pregunta 4: ¿Qué lecciones se pueden extraer de la implementación de sistemas de Machine Learning en la detección de fraudes en otros países que puedan ser aplicables a Colombia?

Lecciones aprendidas en otros países	Aplicación en Colombia
En países como Estados Unidos y Reino Unido, la calidad y variedad de los datos es fundamental para que los modelos de ML detecten patrones de fraude con precisión (Ngai et al., 2019).	Integración diferentes fuentes de datos (transacciones, historial de usuarios) para mejorar la precisión de los modelos de ML.
En países como Australia y Japón, detectar fraudes en tiempo real permite a las instituciones financieras responder rápidamente y evitar pérdidas (Sharma & Panigrahi, 2022).	Adoptar sistemas de detección en tiempo real ayudaría a reducir los fraudes al permitir acciones inmediatas.
En Alemania y Suecia, educar a los usuarios sobre cómo reconocer y evitar fraudes añade una capa importante de protección (Bello & Komolafe, 2024).	Implementar campañas para educar a los usuarios en prácticas seguras e identificación de fraudes.
En Noruega y Países Bajos, la colaboración entre bancos, empresas y gobierno mejora la	Promover la colaboración entre instituciones financieras y el gobierno

Lecciones aprendidas en otros países	Aplicación en Colombia
capacidad de detectar fraudes a nivel nacional (PwC, 2021).	ayudaría a crear una red de datos para analizar patrones de fraude.

Principales tipos de fraudes financieros en Colombia

En el 2024, se ha presentado un incremento en modalidades de fraude financiero, las cuales son:

1. **Suplantación de identidad**: Consiste el robo de identidad para obtener bienes, servicios o información. El 17% de los ciudadanos ha sido víctima de robo de identidad digital en la primera mitad de 2024 (Infobae, 2024).
2. **Fraude digital en transacciones**: Durante el primer semestre de 2024, cerca del 7% de las transacciones realizadas desde Colombia se identificaron como sospechosas de fraude digital, representando un incremento anual del 43,5% respecto al mismo período de 2023 (Portafolio, 2024).
3. **Fraude en la apertura de cuentas**: Mas del 25% de los fraudes se presentó al abrir cuentas de productos o servicios. El fraude de identidad sintética, donde se utilizan identidades falsas para cometer fraudes, aumentó un 153% entre 2023 y lo corrido de 2024 (TransUnion, 2024).
4. **Ataques a la banca móvil mediante malware**: Los ciberdelincuentes utilizan malware avanzado para acceder a aplicaciones de banca móvil y realizar transacciones fraudulentas (La República, 2024).

Mecanismos de *Machine Learning* utilizados para la detección de fraude financiero en Colombia

Modalidad de fraude	Modelos de *Machine Learning* utilizados	Descripción
Suplantación de identidad	Redes neuronales convolucionales (*CNN*)	Identifican patrones en datos biométricos o imágenes de documentos para verificar identidades (Goodfellow et al., 2016).
	Máquinas de soporte vectorial (*SVM*)	Clasifican transacciones y detectan patrones anómalos en los perfiles de los usuarios (Bansal et al., 2024).
	Modelos de análisis de sentimiento y NLP	Analizan patrones de comportamiento y lenguaje en las interacciones para detectar suplantación (Ngai et al., 2019).
Fraude digital en transacciones	Bosques aleatorios (*Random Forest*)	Distinguen transacciones normales de sospechosas en grandes volúmenes de datos en tiempo real (Bansal et al., 2024).
	Arboles de decisión	Detectan patrones específicos en las transacciones, clasificando entre actividades legítimas y fraudulentas (Bello & Komolafe, 2024).
	Isolation forest	Detectan anomalías en datos no estructurados para identificar transacciones inusuales (Charizanos et al., 2024).
	Aprendizaje no supervisado con clustering (*k-means*)	Agrupan transacciones para detectar comportamientos fuera de lo común sin etiquetas previas (Rodrigues et al., 2022).
Fraude en la apertura de cuentas	Redes neuronales profundas (*DNN*)	Validan información y patrones en perfiles históricos, evitando cuentas con datos falsos (Bansal et al., 2024).
	Autoencoders	Reducen dimensionalidad y resaltan diferencias en los datos proporcionados para evitar fraudes (Bello & Komolafe, 2024).

Modalidad de fraude	Modelos de *Machine Learning* utilizados	Descripción
	Algoritmos de clustering no supervisado	Identifican patrones de normalidad y anomalías en los datos de nuevos clientes (Ngai et al., 2019).
Ataques a banca móvil mediante malware	Modelos basados en secuencias temporales	Analizan el comportamiento del usuario para detectar actividades anómalas en dispositivos móviles (Goodfellow et al., 2016).
	Behavioral biometrics	Analizan patrones de comportamiento como velocidad de tecleo y ubicación para evitar accesos no autorizados (PwC, 2021).
	Redes bayesianas	Evalúan la probabilidad de que una actividad esté relacionada con malware mediante datos históricos (McKinsey & Company, 2021).

Plataformas que ofrecen el servicio de detección de fraude en línea

Algunas de las plataformas que ofrece sus servicios para detección de fraude en línea son:

Plataforma	Enfoque	Modelo *Machine Learning*
SAS Fraud Management	Enfoque basado en el comportamiento de los clientes, el cual se analiza mediante *Machine Learning* desde distintas fuentes para identificar incoherencias en tiempo real.	Regresión logística y *Random Forest*
Kount	Enfoque basado en inteligencia artificial y redes neuronales para identificar y prevenir fraudes en transacciones en línea	Arboles de decisión y bosques aleatorios para clasificación, ajuste de los modelos mediante redes neuronales profundas y *Machine Learning* adaptativo

Darktrace	Conocida por sus servicios en ciberseguridad, incluye capacidades de detección de fraudes financieros.	Usa *clustering* y detección de anomalías, *LSTM* y detección de patrones de comportamiento.
Featurespace	Detecta anomalías en tiempo real gracias al uso de la inteligencia artificial, lo que permite la identificación temprana de transacciones fraudulentas.	Aprendizaje no supervisado, detección de anomalías y aprendizaje adaptativo para mejora continua. Utiliza bayes para construir patrones de comportamiento.
FICO Falcon Fraud Manager	Enfoque de análisis de pagos digitales en tiempo real para prevención de fraude	Aprendizaje supervisado, redes neuronales y árboles de decisión y modelos adaptativos
Feedzai	Enfoque en prevención de fraude para pagos financieros	*Random Forest*, redes neuronales profundas y *SVM*

Retos en la implementación

Los principales retos para implementar mecanismos de *Machine Learning* en la detección de fraudes son:

1. Calidad y disponibilidad de datos: Los modelos de ML dependen de datos limpios, representativos y equilibrados. Si los datos están sesgados o tienen errores, el modelo puede no detectar fraudes con precisión, y en contextos de desbalance de clases (fraude vs. transacciones legítimas), el modelo puede no identificar adecuadamente las actividades fraudulentas.
2. Adaptación y evolución del fraude: Los fraudes cambian constantemente, lo que significa que los modelos deben ser actualizados y entrenados continuamente con nuevos datos para mantenerse relevantes y efectivos frente a tácticas de fraude emergentes.
3. Interpretabilidad de los modelos: Algunos modelos avanzados, como las redes neuronales profundas, son difíciles de interpretar. Esto es problemático cuando es necesario explicar las decisiones del modelo (como una alerta de fraude) a los analistas o autoridades, lo que dificulta la validación y confianza en los resultados.

4. Privacidad y cumplimiento normativo: El uso de datos personales y financieros para entrenar modelos de ML debe cumplir con regulaciones de privacidad (como GDPR). Garantizar que los datos sensibles se manejen correctamente sin infringir leyes es un reto importante para la implementación.

Análisis y discusión de los resultados

La aplicación de modelos de *Machine Learning* en la detección de fraude financiero se ha consolidado como una herramienta muy efectiva para identificar patrones complejos y transacciones sospechosas que los métodos tradicionales no logran captar. Estos modelos, especialmente los basados en aprendizaje profundo y los enfoques híbridos, no solo han mejorado la precisión en la detección de fraudes, sino que también han agilizado la respuesta, permitiendo una detección en tiempo real.

Aun así, estos métodos enfrentan desafíos importantes en su implementación práctica. La calidad y disponibilidad de datos son claves para obtener buenos resultados; la efectividad de los modelos depende de que los datos estén balanceados y reflejen adecuadamente los patrones de fraude actuales. Además, aunque estos modelos aumentan la precisión, algunos pueden ser difíciles de interpretar, lo que complica justificar ciertas decisiones ante auditores o reguladores.

Finalmente, es esencial que estos modelos se adapten rápidamente a nuevas tácticas de fraude en un entorno financiero que evoluciona constantemente. Sin embargo, es importante considerar tanto los costos operativos como el cumplimiento de las normativas de privacidad para asegurar que el uso de estas técnicas sea eficiente y ético.

Machine Learning es un campo extremadamente vasto y en constante evolución, lo que deja abiertas muchas posibilidades para explorar nuevos modelos y enfoques que permitan mejorar la detección de fraudes. En el ámbito de las transacciones en línea, los sistemas deben responder en tiempo real para ser efectivos; por ello, los modelos que pueden procesar secuencias de datos y reconocer patrones temporales, como las Redes de Memoria a Largo y Corto Plazo (LSTM, por sus siglas en inglés), han emergido como una de las opciones más

prometedoras. Estas redes, que son capaces de manejar datos en secuencia y aprender de dependencias temporales, son particularmente útiles en transacciones en línea, ya que pueden identificar patrones de comportamiento en cuestión de milisegundos, permitiendo una respuesta rápida y precisa frente a actividades sospechosas.

Sin embargo, es fundamental reconocer que muchos de los estudios disponibles actualmente son ensayos realizados en condiciones de laboratorio con bases de datos limitadas. Estos estudios, aunque valiosos para probar la eficacia de ciertos modelos en escenarios controlados, pueden volverse obsoletos rápidamente al ser aplicados en el mundo real. Los métodos que no se adaptan dinámicamente a nuevos patrones de fraude pueden perder su efectividad, especialmente dado que los defraudadores están en constante innovación. Es por esto que las plataformas líderes en detección de fraude no solo aplican un único modelo estático, sino que integran modelos adaptativos, capaces de ajustar sus algoritmos automáticamente en función del comportamiento cambiante de los delincuentes. Estos modelos permiten que el sistema evolucione y mantenga su relevancia, mejorando la precisión y minimizando falsos positivos.

La mayoría de los estudios relevantes y de alto impacto (Tabla 2) se sustentan en el análisis de datos reales y públicos, lo cual subraya la importancia de disponer de información abierta y accesible para avanzar en la investigación y el desarrollo de modelos más eficaces. En el contexto colombiano, esto representa un desafío significativo: el sector bancario suele ser muy reservado con los datos sobre fraude, en gran medida por temor a que la divulgación de estas incidencias afecte negativamente su reputación y la confianza de sus clientes. Esta reticencia a compartir información se convierte en el *talón de Aquiles* para la innovación en detección de fraudes, ya que limita la posibilidad de crear modelos realmente efectivos en el contexto local. Sin base de datos representativas, cualquier modelo implementado podría carecer de la sensibilidad y adaptabilidad necesaria para responder a las particularidades del fraude en el mercado colombiano. Facilitar un acceso seguro a datos históricos y anónimos sobre fraude podría, sin duda, acelerar el desarrollo de modelos específicos y mejorar la capacidad de respuesta de las instituciones financieras del país. Este aspecto debería ser fácil de conseguir pues los bancos en general gestionan su información de forma digital, lo que facilita la

consolidación de la información, se trataría más bien de un proceso de *anonimización* que permita respaldar estudios y nuevos productos sin señalar a la entidad o a las personas.

Conclusiones

La revisión sistemática de la literatura permitió identificar que el uso de *Machine Learning* en la detección de fraude en tiempo real es un campo de investigación activo y en constante evolución. Los métodos avanzados, como redes neuronales profundas y algoritmos híbridos, destacan en la identificación de patrones complejos que los métodos tradicionales no logran captar, especialmente en un entorno donde los esquemas de fraude evolucionan rápidamente. Este estado del arte indica una tendencia creciente hacia el uso de técnicas más sofisticadas que mejoran la precisión y capacidad de respuesta de los sistemas de detección.

Al comparar los métodos de *Machine Learning* con los enfoques tradicionales, los modelos de aprendizaje profundo y los algoritmos supervisados y no supervisados ofrecen mejoras significativas en términos de precisión y velocidad de respuesta. Los enfoques tradicionales, basados en reglas fijas, requieren actualizaciones constantes y muestran limitaciones para adaptarse a nuevas tácticas de fraude. En contraste, los modelos de *Machine Learning* no solo automatizan la detección, sino que también facilitan una respuesta en tiempo real, lo cual es crítico en el contexto financiero actual.

El análisis de experiencias y estudios de caso de otros países muestra que la implementación de *Machine Learning* en la detección de fraudes ha tenido éxito, especialmente en países con acceso a datos de calidad y recursos computacionales avanzados. Sin embargo, estas experiencias resaltan la necesidad de contar con infraestructuras tecnológicas sólidas y políticas de privacidad adecuadas. Para Colombia, adaptar estas lecciones implica fortalecer el acceso a datos relevantes, considerar la inversión en infraestructura y adoptar un enfoque flexible que permita actualizar y ajustar los modelos continuamente para responder a las tácticas cambiantes de los defraudadores.

En conclusión, aunque el uso de *Machine Learning* en la detección de fraudes aporta claras ventajas sobre los métodos tradicionales, la implementación en Colombia debe enfrentar retos

de calidad de datos, costos operativos y cumplimiento regulatorio. Con el adecuado ajuste a estas condiciones, los modelos de *Machine Learning* tienen el potencial de mejorar significativamente la seguridad y eficiencia en la detección de fraude financiero en el país.

Referencias

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit Card Fraud Detection Using State-of-the-Art *Machine Learning* and Deep Learning Algorithms. IEEE Access, 10, 39700–39715. https://doi.org/10.1109/access.2022.3166891

Alghofaili, Y., Albattah, A., & Rassam, M. A. (2020). A Financial Fraud Detection Model Based on LSTM Deep Learning Technique. Journal of Applied Security Research, 15(4), 498–516. https://doi.org/10.1080/19361610.2020.1815491

Asmar, S. (2023, abril). Provenir se une a Colombia Fintech y le pone IA y *Machine Learning* a las transacciones. La República. Recuperado el 29 de septiembre de 2024, de Provenir se une a Colombia Fintech y le pone IA y *Machine Learning* a las transacciones.

Asobancaria. (2022). Informe de riesgos financieros 2022. Asociación Bancaria y de Entidades Financieras de Colombia. Recuperado de https://www.asobancaria.com/informes/

Aziz, R. M., Baluch, M. F., Patel, S., & Kumar, P. (2022). A *Machine Learning* based Approach to Detect the Ethereum Fraud Transactions with Limited Attributes. Karbala International Journal of Modern Science, 8(2), 139–151. https://doi.org/10.33640/2405-609x.3229

BBVA México. (n.d.). ¿Qué es una transacción financiera? BBVA México. https://www.bbva.mx/educacion-financiera/t/transaccion_financiera.html

Banco de la República. (2023). Reporte de estabilidad financiera. Recuperado el 19 de agosto de 2024, de https://www.banrep.gov.co/es/reporte-estabilidad-financiera

Bansal, A., Garg, H., Herr, D., Obert, B., Rosenkranz, M., Abbassi, H., Mendili, S. E. L., & Gahi, Y. (2024). Digital banking fortification: A real-time isolation forest architecture for detecting online transaction fraud. Engineering Research Express, 6(2), 025214. https://doi.org/10.1088/2631-8695/AD4958

Bello, O. A., & Komolafe, O. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications, challenges, and opportunities. Computer Science & IT Research

Journal, 5(6), 1505-1520. Recuperado el 19 de agosto de 2024, de https://doi.org/10.51594/csitrj.v5i6.1252

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235–255. https://doi.org/10.1214/ss/1042727940

Can, B., Yavuz, A. G., Karsligil, E. M., & Guvensan, M. A. (2020). A Closer Look Into the Characteristics of Fraudulent Card Transactions. IEEE Access, 8, 166095–166109. https://doi.org/10.1109/access.2020.3022315

Charizanos, G., Demirhan, H., & İçen, D. (2024). An online fuzzy fraud detection framework for credit card transactions. Expert Systems with Applications, 252, 124127. https://doi.org/10.1016/J.ESWA.2024.124127

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2022). Graph Neural Network for Fraud Detection via Spatial-Temporal Attention. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3800–3813. https://doi.org/10.1109/tkde.2020.3025588

Dalal, S., Seth, B., Radulescu, M., Secara, C., & Tolea, C. (2022). Predicting Fraud in Financial Payment Services through Optimized Hyper -Parameter-Tuned XGBoost Model. Mathematics, 10(24), 4679. https://doi.org/10.3390/math10244679

Deloitte. (2022). The future of anti-money laundering in the financial services industry. Recuperado el 19 de agosto de 2024, de https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-fsi-future-of-aml-in-fs.pdf

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A Neural Network Ensemble with Feature Engineering for Improved Credit Card Fraud Detection. IEEE Access, 10, 16400–16407. https://doi.org/10.1109/access.2022.3148298

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance Evaluation of *Machine Learning* Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost. IEEE Access, 9, 165286–165294. https://doi.org/10.1109/access.2021.3134330

Infobae. (2024, julio 12). *Top cinco de fraudes financieros que más afectan el bolsillo de los colombianos*. Recuperado de https://www.infobae.com/colombia/2024/07/12/top-cinco-de-fraudes-financieros-que-mas-afectan-el-bolsillo-de-los-colombianos/

Jovanovic, D., Antonijevic, M., Stankovic, M., Zivkovic, M., Tanaskovic, M., & Bacanin, N. (2022). Tuning *Machine Learning* Models Using a Group Search Firefly Algorithm for Credit Card Fraud Detection. Mathematics, 10(13), 2272. https://doi.org/10.3390/math10132272

Kanika et al, "A Survey of Deep Learning based Online Transactions Fraud Detection Systems," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 130-136, doi: 10.1109/ICIEM48762.2020.9160200.

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2008). Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology, 51(1), 7-15. https://doi.org/10.1016/j.infsof.2008.09.009

KPMG. (2023). Global Banking Fraud Survey. Recuperado el 19 de agosto de 2024, de https://home.kpmg/xx/en/home/insights/2023/05/global-banking-fraud-survey.html

La República. (2024, enero 15). *8 amenazas financieras para 2024: Fraudes con IA y ataques a la banca móvil*. Recuperado de https://www.larepublica.net/noticia/8-amenazas-financieras-para-2024-fraudes-con-ia-y-ataques-a-la-banca-movil

McKinsey & Company. (2021). AI-bank of the future: Can banks meet the AI challenge? Recuperado el 19 de agosto de 2024, de https://www.mckinsey.com/industries/financial-services/our-insights/ai-bank-of-the-future-can-banks-meet-the-ai-challenge

Mirkovic, J., & Yang, Y. (2023). *Machine Learning* in fraud detection: Current challenges and future research directions. Journal of Financial Crime, 30(1), 58-75. Recuperado el 19 de agosto de 2024, de https://doi.org/10.1108/JFC-07-2022-0112

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2019). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569. https://doi.org/10.1016/j.dss.2010.08.006

Oficina para la Protección Financiera del Consumidor. (n.d.). ¿Qué es una transferencia electrónica? Oficina para la Protección Financiera del Consumidor. https://www.consumerfinance.gov/es/obtener-respuestas/que-es-una-transferencia-electronica-es-1163/

Oluwabusayo Adijat Bello, & Komolafe Olufemi. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications, challenges and opportunities. Computer Science & IT Research Journal, 5(6), 1505-1520. https://doi.org/10.51594/csitrj.v5i6.1252

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*, 34(3), 279–310. https://doi.org/10.1007/s10462-010-9179-y

Portafolio. (2024, abril 17). *Intentos de fraude digital en Colombia crecen el 43,5%*. Recuperado de https://www.portafolio.co/tecnologia/intentos-de-fraude-digital-en-colombia-crecen-el-43-5-616150

PwC. (2021). Global Economic Crime and Fraud Survey 2021. PricewaterhouseCoopers. Recuperado el 19 de agosto de 2024, de https://www.pwc.com/gx/en/services/forensics/economic-crime-survey.html

Rodrigues, V. F., Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., da Costa, C. A., Barbosa, J. L. V., Antunes, R. S., Scorsatto, R., & Arcot, T. (2022). Fraud detection and prevention in e-commerce: A systematic literature review. Electronic Commerce Research and Applications, 56, 101207. https://doi.org/10.1016/J.ELERAP.2022.101207

Santander. (2020). Annual Report 2020. Santander Group. Recuperado el 19 de agosto de 2024, de https://www.santander.com/csgs/Satellite/CFWCSancomQP01/en_GB/Corporate/Shareholders-and-Investors.html

SAS Institute. (2022). Fraude en el sector financiero: Cómo combatirlo con analítica avanzada. Recuperado el 19 de agosto de 2024, de https://www.sas.com/es_co/insights/articles/risk-fraud/fraude-en-el-sector-financiero.html

Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2021). An intelligent payment card fraud detection system. Annals of Operations Research, 334(1–3), 445–467. https://doi.org/10.1007/s10479-021-04149-2

Sharma, A., & Panigrahi, B. K. (2022). A hybrid approach for detecting financial frauds using *Machine Learning* techniques. Computers & Security, 105, 102222. https://doi.org/10.1016/j.cose.2021.102222

Superintendencia Financiera de Colombia. (2023). Informe de crecimiento del sector financiero 2023. Recuperado el 19 de agosto de 2024, de https://www.superfinanciera.gov.co/informes/

Superintendencia Financiera de Colombia. (2024). Informe de operaciones: Segundo trimestre de 2024. Recuperado el 02/11/2024 de https://www.superfinanciera.gov.co/publicaciones/10115298/informe-de-operaciones-segundo-trimestre-de-2024/#:~:text=En%20el%20segundo%20trimestre%20de,y%201.660%20millones%20no%20monetarias).

Towards Data Science. (2020, mayo 6). *Fraud Detection with Machine Learning*. Towards Data Science. Recuperado de https://towardsdatascience.com

Vanini, P., Rossi, S., Zvizdic, E., & Domenig, T. (2023). Online payment fraud: from anomaly detection to risk management. Financial Innovation, 9(1). https://doi.org/10.1186/s40854-023-00470-w

Villamil Arcos, C. (2022). Selección de una técnica de aprendizaje de máquina para la detección de Fraude Financiero Digital enfocado a transacciones no autorizadas o consentidas. Universidad Nacional de Colombia

Wang, J., Xu, Y., & Zhou, Z. (2022). Enhancing financial fraud detection with adversarial learning. Neurocomputing, 473, 35-45.

Wang, C., Wang, C., Zhu, H., & Cui, J. (2021). LAW: Learning Automatic Windows for Online Payment Fraud Detection. IEEE Transactions on Dependable and Secure Computing, 1–1. https://doi.org/10.1109/tdsc.2020.3037784

Wang, H., Wang, W., Liu, Y., & Alidaee, B. (2022). Integrating *Machine Learning* Algorithms with Quantum Annealing Solvers for Online Fraud Detection. IEEE Access, 10, 75908–75917. https://doi.org/10.1109/access.2022.3190897

Zhang, Y., Li, X., & Wang, Q. (2022). Application of multi-criteria decision analysis in renewable energy investment: A comprehensive review. Applied Sciences, 12(19), 9637. https://www.mdpi.com/2076-3417/12/19/9637.