COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MÁSTER EN TECNOLOGÍAS FINANCIERAS:
PAGOS Y BANCA DIGITAL

TRABAJO FIN DE MÁSTER

ESTUDIO DE MODELOS DE DETECCIÓN DE MULEROS EN LA BANCA DIGITAL: ANÁLISIS DE LIMITACIONES Y NUEVAS ESTRATEGIAS

Autor: **Matteo Filippo Massarelli**

Director: **Antolín Martínez Martínez**

Madrid

Julio de 2025

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

Estudio de modelos de detección de muleros en la banca digital: análisis de limitaciones y nuevas estrategias.

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2024/25 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Matteo Filippo Massarelli Fecha: 17 / 07 / 2025

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

ANTOLIN| MARTINEZ| MARTINEZ

Firmado digitalmente por ANTOLIN|MARTINEZ| MARTINEZ
Fecha: 2025.07.17 10:35:12 +02'00'

Fdo.: Antolín Martínez Martínez Fecha: ……/ ……/ ……

COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MÁSTER EN TECNOLOGÍAS FINANCIERAS:
PAGOS Y BANCA DIGITAL

TRABAJO FIN DE MÁSTER

ESTUDIO DE MODELOS DE DETECCIÓN DE MULEROS EN LA BANCA DIGITAL: ANÁLISIS DE LIMITACIONES Y NUEVAS ESTRATEGIAS

Autor: **Matteo Filippo Massarelli**

Director: **Antolín Martínez Martínez**

Madrid

Julio de 2025

Agradecimientos

Dedico estas líneas a quienes me han apoyado en este camino. Al profesorado del máster, mi sincero agradecimiento por su dedicación y guía, que han sido una pieza clave en mi formación. Quiero agradecer también al equipo de Business Analytics de NTT Data la confianza que me brindaron; su oportunidad fue fundamental para mi crecimiento profesional y para poder poner en práctica lo aprendido.

En lo personal, mi mayor gratitud es para mi familia. Su apoyo incondicional y su paciencia, durante los momentos de mayor dedicación, han sido el pilar que me ha permitido completar este viaje.

ESTUDIO DE MODELOS DE DETECCIÓN DE MULEROS EN LA BANCA DIGITAL: ANÁLISIS DE LIMITACIONES Y NUEVAS ESTRATEGIAS

Autor: Massarelli, Matteo Filippo.
Director: Martínez Martínez, Antolín.
Entidad Colaboradora: NTT DATA y ICAI - Universidad Pontificia Comillas

RESUMEN DEL PROYECTO

La digitalización del sector financiero, si bien ha democratizado el acceso a servicios, ha generado una nueva superficie de ataque para el crimen organizado, donde las cuentas mula se han convertido en un eslabón estructural. Estas cuentas son el vehículo indispensable para ofuscar el rastro del dinero, conectando así el fraude con el blanqueo de capitales. Este trabajo diagnostica que los modelos de detección vigentes sufren limitaciones sistémicas: operan sobre la falta de integración de datos y con una visión fragmentada del cliente, son vulnerables a la deriva conceptual, el fenómeno por el cual los patrones de fraude cambian con el tiempo, dada su naturaleza reactiva y su eficacia se ve mermada por el desbalance de clases y una ceguera estructural ante las redes criminales.

Frente a esta problemática, el objetivo principal es proponer un cambio de paradigma: el diseño de un modelo de detección holístico y proactivo, enfocado en el análisis integral del ciclo de vida del cliente y se potencia con el análisis de grafos para desvelar las redes delictivas. El diseño de la metodología parte de una revisión crítica de los sistemas actuales para, a partir de sus debilidades, formular una arquitectura alternativa que descompone la actividad del mulero en fases (alta, operativa, desconexión/huida) y utiliza el análisis de grafos para identificar patrones de colusión.

El alcance del trabajo es teórico y estratégico, sin una implementación técnica directa. Su validez, no obstante, se fundamenta en una rigurosa revisión de la literatura y de la propia experiencia práctica acumulada en NTT DATA. Allí, distintos proyectos ya implementan con éxito sus elementos, lo cual constituye una prueba que confirma tanto su pertinencia como su notable potencial futuro. En definitiva, este trabajo aporta un marco estratégico para que las entidades financieras transiten desde una postura de contención de daños

hacia una de anticipación y desarticulación de amenazas, fortaleciendo así la integridad y la confianza en la banca digital.

Palabras clave: Fraude financiero, cuentas mula, banca digital, machine learning, análisis de grafos, AML, detección de anomalías, GDPR, PSD2.

1. Introducción

La transformación digital de la banca ha democratizado el acceso financiero, pero a costa de crear una sofisticada superficie de ataque para el crimen organizado. Dentro de ella, las cuentas mula actúan como el eslabón clave que conecta fraude y blanqueo, socavando la integridad del sistema. Los sistemas de defensa actuales se revelan insuficientes. Su naturaleza es reactiva, operan con datos fragmentados y analizan cuentas de forma aislada, una ceguera que impide detectar redes coordinadas. Para subsanar estas carencias, este trabajo presenta el diseño de un modelo proactivo, cuyo enfoque, centrado en el ciclo de vida del cliente y potenciado por análisis de grafos, busca construir una defensa resiliente y anticipatoria.

2. Definición del proyecto

Este proyecto diseña conceptualmente una arquitectura conceptual para la detección de cuentas mula que se define por ser holística, proactiva e integrada. El modelo se fundamenta en el análisis del ciclo de vida del cliente, identificando patrones anómalos desde el origen. Se potencia, a su vez, con el análisis de grafos para desvelar redes criminales ocultas a los sistemas actuales. El alcance se mantiene en el plano teórico y estratégico; no se contempla una implementación técnica dadas las barreras regulatorias y de confidencialidad (GDPR) del sector, sino que el foco es formular una propuesta pertinente para las necesidades de las entidades bancarias.

3. Descripción del modelo/sistema/herramienta

Análisis del propio ciclo de vida y la aplicación de análisis de grafos.

4. Resultados

Como estudio teórico, este trabajo culmina en el diseño de una arquitectura conceptual para un modelo de detección holístico, proactivo y relacional. A diferencia de los sistemas vigentes, este modelo logra una visión integral del cliente al conectar datos de fuentes tradicionalmente aisladas e interviene proactivamente desde el alta del cliente. Proporciona, además, una visión completa de las redes criminales con análisis de grafos. Se integra un bucle de aprendizaje continuo, crucial para combatir la constante evolución de las tácticas de fraude, y un marco de gobernanza que, usando técnicas de IA Explicable (XAI), asegura la interpretabilidad de las decisiones, garantizando su transparencia y sostenibilidad.

5. Conclusiones

Los paradigmas actuales de detección de cuentas mula son, por su propio diseño, reactivos y fragmentados. Operan con una "visión de túnel transaccional", actuando solo cuando el daño ya está consumado. Frente a esto, la propuesta central del trabajo es un cambio de paradigma: un modelo integrado que se basa en el ciclo de vida del cliente y se potencia con análisis de grafos. Este enfoque permite una evolución clave, pasando de clasificar eventos aislados a interpretar narrativas criminales; del marcaje de nodos individuales a la desarticulación de redes enteras. Se ofrece así una hoja de ruta conceptual para que las instituciones financieras transiten desde la contención de daños hacia la anticipación de amenazas, fortaleciendo la integridad del ecosistema financiero digital.

STUDY OF MONEY MULE DETECTION MODELS IN DIGITAL BANKING: ANALYSIS OF LIMITATIONS AND NEW STRATEGIES

Author: Massarelli, Matteo Filippo.
Supervisor: Martínez Martínez, Antolín.
Collaborating Entity: NTT DATA y ICAI - Universidad Pontificia Comillas

ABSTRACT

The digitalization of the financial sector, while having democratized access to services, has generated a new attack surface for organized crime, where mule accounts have become a structural link. These accounts are the indispensable vehicle for obfuscating the money trail, connecting fraud with money laundering. This work diagnoses how current detection models suffer from systemic limitations: they operate on information silos with a fragmented view of the client, are vulnerable to conceptual drift—the phenomenon by which fraud patterns change over time—their reactive nature, and their effectiveness is undermined by class imbalance and a structural blindness to criminal networks.

Faced with this problem, the main objective is to propose a paradigm shift: the design of a holistic and proactive detection model, focused on the integral analysis of the client's life cycle and enhanced with graph analysis to uncover criminal networks. The design of the methodology starts from a critical review of the current systems to, based on their weaknesses, formulate an alternative architecture that breaks down the mule's activity into phases (registration, operation, disconnection/escape) and uses graph analysis to identify patterns of collusion.

The scope of the work is theoretical and strategic, without direct technical implementation. Its validity, however, is based on a rigorous review of the literature and the practical experience accumulated at NTT DATA. There, various projects already successfully implement its elements, which constitutes proof that confirms its relevance and its notable future potential. Ultimately, this work provides a strategic framework for financial institutions to transition from a posture of damage containment to one of anticipation and dismantling of threats, thereby strengthening integrity and trust in digital banking.

Keywords: Financial fraud, money mule accounts, digital banking, machine learning, graph analysis, AML, anomaly detection, GDPR, PSD2.

1. Introduction

The digital transformation of banking has democratized financial access, but at the cost of creating a sophisticated attack surface for organized crime. Within it, money mule accounts act as the key link connecting fraud and laundering, undermining the integrity of the system. Current defense systems are proving insufficient. Their nature is reactive, they operate with fragmented data, and they analyze accounts in isolation—a blindness that prevents the detection of coordinated networks. To address these shortcomings, this work articulates a proactive conceptual model whose approach, centered on the customer lifecycle and enhanced by graph analysis, seeks to build a more resilient and anticipatory defense.

2. Project Definition

This project designs a conceptual architecture for money mule detection that is defined by being holistic, proactive, and integrated. The model is based on the analysis of the customer lifecycle, identifying anomalous patterns from their origin. It is, in turn, enhanced by graph analysis to uncover criminal networks hidden from current systems. The scope remains on a theoretical and strategic level; a technical implementation is not contemplated given the regulatory and confidentiality barriers (GDPR) of the sector, with the focus instead being on formulating a proposal relevant to the industry's needs.

3. Model/System/Tool Description

Analysis of the lifecycle itself and the application of graph analysis.

4. Results

As a theoretical study, this work culminates in the design of a conceptual architecture for a holistic, proactive, and relational detection model. Unlike current systems, this model achieves a comprehensive view of the client by connecting data from traditionally isolated sources and intervenes proactively from client onboarding. Furthermore, it provides a complete view of criminal networks through graph analysis. A continuous learning loop is integrated, crucial for combating the constant evolution of fraud tactics, along with a

governance framework that, using Explainable AI (XAI) techniques, ensures the interpretability of decisions, guaranteeing their transparency and sustainability.

5. Conclusions

Current paradigms for money mule detection are, by their very design, reactive and fragmented. They operate with a "transactional tunnel vision," acting only after the damage has been done. In response, the central proposal of this work is a paradigm shift: an integrated model based on the customer lifecycle and enhanced by graph analysis. This approach allows for a key evolution, moving from classifying isolated events to interpreting criminal narratives; from flagging individual nodes to dismantling entire networks. It thus offers a conceptual roadmap for financial institutions to transition from damage containment to threat anticipation, strengthening the integrity of the digital financial ecosystem.

Índice de la memoria

Índice de figuras

Índice de tablas

CAPÍTULO 1. INTRODUCCIÓN

La digitalización ha reconfigurado los cimientos del sector financiero en la última década, trascendiendo la mera optimización de procesos para transformar el modelo de negocio bancario. Esta transición desde una operativa centrada en la sucursal física hacia un ecosistema digital, hiperconectado y operativo en tiempo real, ha democratizado el acceso a servicios, generando ganancias exponenciales en eficiencia y experiencia de cliente. Innovaciones como la banca abierta (Open Banking) o los pagos instantáneos han desdibujado las fronteras tradicionales, catalizando un mercado más competitivo. Esta transformación, si bien indispensable, ha abierto simultáneamente una nueva superficie de ataque para la delincuencia financiera, de una escala y sofisticación antes inimaginables.

Las mismas tecnologías que permiten a un cliente abrir una cuenta en minutos o ejecutar transferencias transfronterizas al instante, han sido instrumentalizadas por organizaciones criminales para operar con una velocidad y un anonimato que desbordan las estructuras de control convencionales. El crimen financiero ha mutado desde fraudes locales hacia un fenómeno global y tecnológicamente avanzado. En este paradigma, las barreras físicas han ido desapareciendo, permitiendo a las redes delictivas explotar las vulnerabilidades de un ecosistema fragmentado por jurisdicciones. La capacidad de ejecutar cadenas de transacciones complejas en segundos no solo ha optimizado la operativa legítima, sino que ha perfeccionado la anatomía del blanqueo de capitales.

Esta nueva arquitectura delictiva se centra en una figura clave, un eslabón indispensable para conectar el fraude con el blanqueo: la cuenta mula. Dicho instrumento, en apariencia una cuenta ordinaria, funciona como el vehículo intermediario para recibir, estratificar y legitimar fondos ilícitos. La expansión de cuentas mula se ha convertido en un pilar del crimen organizado que amenaza directamente la integridad estructural de la banca. Las organizaciones criminales han industrializado el reclutamiento de muleros, apalancándose en las redes sociales y la ingeniería social para captar a individuos, a menudo vulnerables, convirtiéndolos en la primera línea, consciente o inconsciente, de sus operaciones. La facilidad para abrir estas cuentas, la velocidad de los flujos y la dificultad del rastreo han convertido a las redes de mulas en el principal catalizador del

fraude. Abordar este fenómeno trasciende la mera gestión del riesgo, sino un imperativo para proteger la reputación institucional. Una entidad puede convertirse en vehículo involuntario de delitos que dañan a sus clientes y, en última instancia, amenazan la integridad del sistema financiero en su conjunto.

1.1. FORMULACIÓN DEL PROBLEMA: CUENTAS MULA Y RETOS DE DETECCIÓN

Una cuenta mula es, en esencia, una cuenta de depósito perteneciente a un titular legítimo que es utilizada, con o sin su consentimiento, como un conducto para fondos delictivos, con el fin de ofuscar su origen. El titular, denominado "mulero" (money mule), actúa como un intermediario que permite a las organizaciones criminales distanciarse del flujo financiero. La comprensión de este fenómeno exige una taxonomía precisa de los perfiles implicados.

El ciclo de vida del mulero sigue un patrón que, sin embargo, los sistemas actuales luchan por identificar de forma holística. Comienza con una fase de reclutamiento y alta (onboarding), donde se explota la eficiencia digital para crear cuentas con perfiles "limpios". Le sigue una fase de explotación activa, caracterizada por la recepción súbita de fondos y su inmediata dispersión mediante múltiples transacciones, a menudo en un patrón de "nivelación" que deja el saldo próximo a cero. Finalmente, llega la fase de desconexión, en la que la cuenta es abandonada una vez "quemada", dejando al titular expuesto a graves consecuencias legales.

La detección eficaz de estas operativas se enfrenta a desafíos formidables. El primero es la fragmentación de los datos: la información del cliente reside en silos inconexos, es decir, en repositorios de datos que operan de forma aislada y sin comunicación entre si, impidiendo una visión 360 grados. El segundo es la adaptabilidad del adversario: las redes criminales aprenden y modifican sus tácticas, generando un constante concept drift que degrada los modelos predictivos. En tercer lugar, el profundo desbalance de clases, el fraude es una ínfima minoría, dificulta el entrenamiento de algoritmos de aprendizaje supervisado. Finalmente, y quizás el reto más complejo, es la opacidad relacional: los modelos tradicionales analizan cuentas de forma aislada, siendo incapaces de detectar las

estructuras de red coordinadas que son la verdadera firma del crimen organizado. Identificar una cuenta mula individual es útil, pero desarticular la red que lo conecta con otros es un objetivo inalcanzable con la actual visión de túnel transaccional.

1.2. OBJETIVOS DEL ESTUDIO Y DELIMITACIÓN DEL ALCANCE

Este Trabajo Fin de Máster surge como respuesta a las limitaciones sistémicas expuestas, con la ambición de proponer un cambio de paradigma en la detección de cuentas mula. El objetivo principal de este estudio es diseñar conceptualmente una arquitectura de detección alternativa, de carácter holístico y proactivo, que se articule en torno al análisis integral del ciclo de vida del cliente. En lugar de centrarse en la clasificación de transacciones discretas, este enfoque busca identificar patrones de comportamiento anómalos a lo largo de la relación del cliente con la entidad.

Para alcanzar este fin, se persigue realizar un diagnóstico crítico de las debilidades de los modelos de detección vigentes, desde el tratamiento del dato hasta la lógica de los motores de decisión. A partir de ahí, se desarrollará un marco teórico que descomponga el ciclo de vida del mulero en fases clave, identificando para cada una las señales de riesgo y los patrones conductuales. Posteriormente, se explorará el potencial del análisis de grafos como herramienta para superar la visión fragmentada y desvelar redes criminales. Finalmente, se propondrá una arquitectura funcional que combine el enfoque de ciclo de vida con estas técnicas avanzadas, sentando las bases para una nueva generación de sistemas de defensa más resilientes.

Diseñar e implementar un sistema de detección de cuentas mula basado en analítica avanzada es una tarea de considerable complejidad. Por esa razón, el alcance de este Trabajo Fin de Máster ha sido delimitado con precisión. El proyecto se enfoca, por tanto, en el diseño de una arquitectura de detección alternativa. Se trata de un estudio eminentemente teórico y estratégico, excluyendo de forma explícita su implementación técnica.

Esta delimitación obedece tanto a la complejidad del proyecto como a las barreras del sector. Las normativas de privacidad, como el GDPR, impiden el acceso a datos reales en un entorno académico. No obstante, la validez del enfoque se basa sobre un doble pilar:

el rigor de la revisión bibliográfica y la experiencia profesional adquirida en NTT DATA. Este anclaje en un ecosistema real de desarrollo antifraude es crucial. Asegura que el modelo trascienda el mero ejercicio abstracto para ser una propuesta alineada con las necesidades de la industria.

El propósito es que este trabajo constituya una pieza esencial dentro de un proyecto más amplio y global. Sienta así las bases que, a futuro, podrían desarrollarse en un entorno real de consultoría tecnológica y en colaboración directa con las entidades bancarias.

1.3. ESTRUCTURA DEL DOCUMENTO

Para guiar al lector a través de la argumentación, este trabajo se ha estructurado en una progresión lógica que avanza desde el diagnóstico del problema hasta la formulación de una solución.

Tras esta introducción, el *Capítulo 2* se adentrará en una comprensión profunda del fenómeno de las cuentas mula, analizando su papel en el ecosistema delictivo, desambiguando conceptos clave como fraude y estafa, y revisando el marco regulatorio.

El *Capítulo 3* se dedicará a un análisis crítico de las limitaciones de los enfoques de detección actuales. Se examinarán las debilidades que se originan tanto en el tratamiento del dato como en los propios motores de decisión, concluyendo que el paradigma vigente es, por diseño, reactivo y fragmentado.

Como respuesta, el *Capítulo 4* presentará el núcleo de la propuesta: el modelo de detección integrado basado en el ciclo de vida del cliente. Se descompondrá el ciclo en sus fases para detallar cómo un análisis contextual y conductual permite una evaluación proactiva del riesgo.

El Capítulo 5 se centra en el análisis de grafos, detallando cómo esta tecnología superan la ceguera relacional de los sistemas tradicionales y permiten pasar de la detección de nodos aislados al desmantelamiento de la infraestructura criminal.

Finalmente, el *Capítulo 6* presentará las conclusiones generales del estudio, sintetizando los hallazgos, recapitulando las ventajas del modelo propuesto y delineando futuras líneas de investigación. En conjunto, esta estructura busca ofrecer una visión completa y

rigurosa, que no solo identifica un problema crítico, sino que articula una solución conceptual robusta e innovadora.

CAPÍTULO 2. COMPRENSIÓN DEL FENÓMENO DE LAS CUENTAS MULAS

2.1 CRIMEN FINANCIERO EN LA ERA DIGITAL: MARCO CONCEPTUAL Y EVOLUCIÓN

El crimen financiero, históricamente asociado a economías sumergidas o fraudes de radio geográfico limitado, ha transmutado en un fenómeno global, tecnológicamente sofisticado y de alcance transnacional que desborda las estructuras de control tradicionales. Las barreras físicas que antes contenían la actividad ilícita se han diluido, lo que permite a las organizaciones criminales operar con una escala y velocidad inimaginables hace apenas una década. La facultad de ejecutar transacciones complejas entre múltiples jurisdicciones en segundos introduce vulnerabilidades sistémicas, explotadas hoy de forma metódica por redes delictivas con una organización y resiliencia crecientes.

Un análisis riguroso del problema de las cuentas mula requiere, por ello, una delimitación precisa de las tipologías delictivas que aquí convergen. Si bien los términos "crimen financiero", "fraude" y "estafa" a menudo se usan como sinónimos en el discurso público, desde una óptica técnica y jurídica sus significados difieren, con implicaciones operativas y de responsabilidad que son cruciales. Entender estas diferencias no constituye un mero ejercicio semántico; es el cimiento para diseñar estrategias de prevención y detección realmente efectivas.

2.1.1. Desambiguación Conceptual: Delimitando el Crimen Financiero, el Fraude y la Estafa

Bajo la rúbrica de crimen financiero se agrupa la categoría más amplia, que abarca un vasto espectro de actividades ilícitas no violentas cuyo objetivo último es la obtención de un beneficio económico indebido mediante la manipulación de sistemas financieros. Organismos como el Grupo de Acción Financiera Internacional (FATF) y la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC) incluyen bajo este concepto delitos diversos como el blanqueo de capitales, la financiación del terrorismo, el fraude fiscal, la corrupción o el abuso de mercado, además del fraude en sus múltiples formas. El nexo común de estas actividades es el uso del engaño, el abuso de confianza o la ocultación para subvertir la integridad de los flujos económicos.

Dentro de este universo, resulta indispensable trazar una distinción técnica fundamental entre los conceptos de fraude y estafa. Esta diferenciación posee implicaciones profundas para la atribución de responsabilidad y el diseño de estrategias de defensa.

El fraude, en su acepción técnica, define una actividad delictiva en la que el perpetrador obtiene un beneficio económico sin que la víctima conozca o autorice la transacción final. La víctima puede ser inducida mediante engaño a revelar información sensible en una fase previa, pero no consiente el acto de disposición patrimonial que consuma el delito. La esencia del fraude yace en que el criminal se apropia de los activos suplantando la identidad de la víctima o explotando sus credenciales ante un tercero, que suele ser una entidad financiera. La institución es, por tanto, engañada para procesar una operación que asume como legítima.

Ejemplos paradigmáticos de fraude son: el fraude de tarjeta no presente (CNP), donde el delincuente emplea datos de una tarjeta robada para compras online o telefónicas; el robo de identidad (Identity Theft), por el que un criminal usa datos personales de la víctima para abrir cuentas o créditos a su nombre, generando deudas que la persona afectada desconoce; o la toma de control de cuentas (Account Takeover - ATO), en la que los perpetradores, tras obtener las credenciales de la banca online, proceden a vaciar las cuentas con transferencias no autorizadas. En cada uno de estos escenarios, el rasgo definitorio es la ausencia de autorización de la víctima en la transacción.

La estafa, en cambio, invierte esta dinámica. En una estafa, el delincuente no opera de forma subrepticia; manipula psicológicamente a la víctima para que esta, de forma voluntaria, realice o autorice una transferencia de fondos a una cuenta controlada por el criminal. La transacción, desde un punto de vista técnico, es autorizada. Este delito es conocido en la industria como “fraude de pago autorizado” (Authorised Push Payment (APP) Fraud). Su éxito no depende solo de la tecnología, sino del dominio de la ingeniería social: el arte de explotar emociones humanas, confianza, miedo, avaricia, afecto, para anular el juicio de la víctima y persuadirla a actuar contra sus propios intereses.

Las tipologías de estafa son diversas y evolucionan constantemente, adaptándose al contexto social y tecnológico. Entre las más prevalentes se encuentran las estafas de inversión, que prometen altos rendimientos con bajo riesgo mediante esquemas Ponzi o piramidales. También destacan las estafas románticas, donde se forja una relación emocional para luego solicitar dinero aduciendo una emergencia, y la suplantación de identidad de organismos. En esta última, los criminales se hacen pasar por entidades de confianza (Hacienda, bancos...) usando llamadas (vishing), SMS (smishing) o emails (phishing) para alertar de un falso problema e instar a un pago inmediato.

La trascendencia de esta distinción conceptual va más allá de lo académico, pues impacta directamente en la atribución de la responsabilidad financiera. En muchos marcos regulatorios, ante un fraude, la entidad financiera debe reembolsar los fondos. Sin embargo, en una estafa de pago autorizado, la situación es más ambigua. Al haber autorizado el cliente la operación, la recuperación del dinero se complica y la responsabilidad suele recaer en la víctima. Esta "brecha de responsabilidad" alimenta un intenso debate regulatorio. En respuesta, jurisdicciones como el Reino Unido han empezado a implementar normativas que obligan a las entidades a compartir la carga del reembolso, reconociendo así su deber de protección.

En este ecosistema emerge la figura de la mula de dinero (money mule): un intermediario que presta su cuenta para recibir y transferir fondos de origen ilícito. Aunque la mula pueda ser captada mediante una estafa, su función es instrumental para el fraude y, de forma inseparable, para el blanqueo de capitales (AML).

2.1.2. Magnitud y Anatomía del Impacto Económico Global

El blanqueo de capitales, por contra, no se ocupa de generar la ganancia inicial, sino de ocultar el origen ilícito de fondos ya existentes para así poder integrarlos en el circuito financiero legítimo. Para ello, se vale de un proceso de estratificación (layering) concebido específicamente para borrar su rastro. Por su parte, la financiación del terrorismo, si bien emplea técnicas de ocultación análogas a las del blanqueo, obedece a una motivación ideológica y no lucrativa, y puede nutrirse de fondos tanto lícitos como ilícitos.

Este complejo proceso de legitimación de capitales se articula comúnmente en tres fases secuenciales, cuyo objetivo es distanciar el dinero de su origen delictivo hasta dotarlo de una apariencia legal. Como se ilustra en la *Figura 1.* a continuación, estas etapas son conocidas como colocación, estratificación e integración.

Figura 1. Etapas principales en las actividades de blanqueo de capitales (Elaboración basada en [1])

La primera fase, la colocación, responde al desafío más inmediato y vulnerable para el blanqueador: la introducción física del dinero en efectivo en el sistema financiero. Este paso suele materializarse a través de depósitos en entidades bancarias, frecuentemente fraccionados en importes menores para eludir los controles de alerta, una técnica conocida como “smurfing”, o mediante la adquisición de instrumentos monetarios como cheques de gerencia.

Una vez que los fondos han penetrado en el sistema, se inicia la etapa de estratificación, diseñada para ofuscar su procedencia y desvincularlos de la actividad criminal que los generó. El dinero se desplaza electrónicamente entre múltiples cuentas, a menudo dispersas en distintas jurisdicciones y paraísos fiscales (offshore), y se emplea en la compraventa de productos de inversión o se canaliza entre empresas pantalla con facturación simulada. El objetivo es tejer un rastro documental tan confuso que la auditoría de los fondos resulte prácticamente inviable.

Finalmente, la fase de integración culmina el ciclo. En esta última etapa, los fondos, ya "limpios", retornan al circuito económico aparentando provenir de fuentes legítimas. El dinero se materializa a través de la compra de activos de alto valor (inmuebles, yates), la inversión en negocios con una fachada legal (restaurantes, constructoras) o la creación de carteras financieras. En este punto, el capital blanqueado es ya indistinguible de cualquier otro de procedencia lícita.

La magnitud de este fenómeno es abrumadora y representa una amenaza sistémica para la integridad económica global. Estimaciones conservadoras sitúan el volumen de capitales blanqueados anualmente entre el 2% y el 5% del Producto Interior Bruto mundial [2]. En el contexto europeo, se calcula que flujos ilícitos por valor de 750.2 mil millones de dólares transitaron por el sistema financiero en 2023, con pérdidas asociadas al fraude que superaron los 103.6 mil millones de dólares en el mismo período [2]. Estas pérdidas por fraude, que afectan tanto a consumidores como a las propias entidades, no se distribuyen de manera uniforme por el continente. Como se puede ver en la Figura 2, existe una concentración significativa en las economías de mayor tamaño, destacando el Reino Unido, Francia y Alemania como los países con las cifras más elevadas [2].

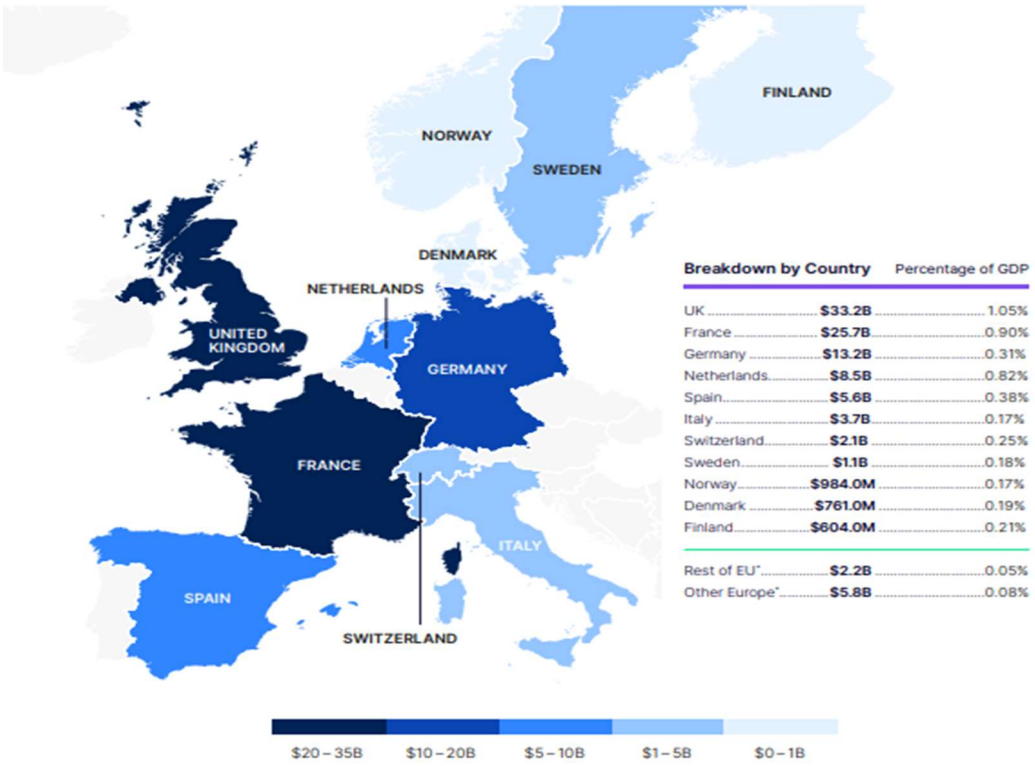

Figura 2. Distribución y Magnitud de las Pérdidas por Fraude en Europa (2023) [2]

Nota. *El mapa ilustra las pérdidas totales por fraude en miles de millones de dólares estadounidenses ($), destacando las jurisdicciones con mayor incidencia. Los datos también se presentan como porcentaje del PIB de cada país*

De forma más específica, se estima que hasta 58.2 mil millones de dólares del blanqueo europeo son directamente atribuibles a la operativa con cuentas mula [2]. Tales cifras no solo dimensionan el problema: revelan una simbiosis crítica. El fraude digital genera los fondos que alimentan las cadenas de blanqueo, en las que la figura del mulero bancario se ha vuelto un eslabón indispensable.

Este diagnóstico no es una mera inferencia analítica. Al contrario, sitúa a las cuentas mula como el nexo indispensable entre fraude y blanqueo, encontrando su validación más contundente en la percepción del riesgo del sector. La vieja visión del mulero como simple cómplice de bajo nivel o problema colateral ha quedado definitivamente atrás. La industria financiera lo reconoce ahora por lo que es: una pieza estructural. Un engranaje fundamental dentro de la maquinaria del crimen organizado contemporáneo que socava la integridad misma del sistema. Las cuentas mula son, en efecto, el vehículo que

materializa la simbiosis crítica del delito financiero digital. Sin ellas, los ingentes fondos del fraude no podrían inyectarse con la velocidad requerida en las cadenas de blanqueo.

La Figura 3 ilustra esta nueva realidad de forma irrefutable. Ciertamente, existen lógicas divergencias regionales en la priorización de amenazas. A pesar de estas diferencias, las cuentas mula emergen como una constante de alta preocupación en todos los tableros. De hecho, en la Unión Europea constituyen la segunda mayor amenaza percibida (37%), mientras que en Reino Unido y los países nórdicos ocupan el tercer puesto (34%). [2]

Resulta profundamente revelador que la amenaza de los muleros se sitúe al mismo nivel de prioridad que conceptos tan graves como crimen organizado o financiación del terrorismo. Esto implica nada menos que un cambio de paradigma en la industria. Ya no se trata de un simple problema de gestión de pérdidas por fraude. Se ha convertido en un vector que, por su escala y función, amenaza directamente la estabilidad, la confianza y la seguridad del ecosistema financiero digital en su conjunto.

Esta elevada y unánime percepción del riesgo constituye la validación definitiva que impulsa este trabajo. Abordar el fenómeno de las cuentas mula ya no es una tarea meramente operativa; se ha transformado en un imperativo estratégico de primer orden. Por ello, la propuesta de un nuevo paradigma de detección trasciende la mera innovación técnica deseable. Se erige como una respuesta necesaria y urgente a uno de los desafíos más críticos que enfrenta hoy la banca moderna.

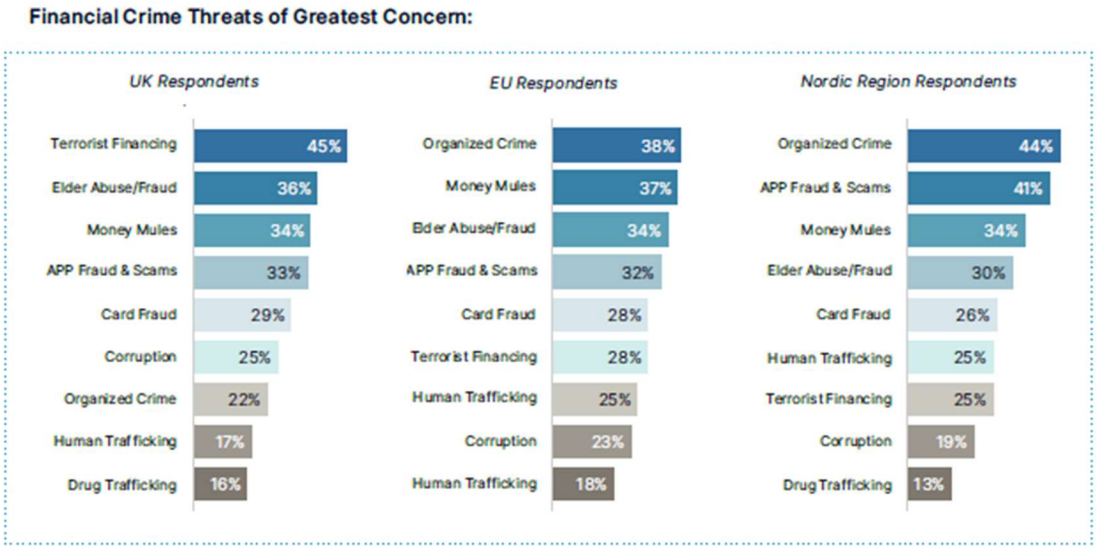

Figura 3. Principales amenazas de delincuencia financiera según la percepción de los profesionales del sector por región (2024) [2]

La digitalización ha acelerado este ciclo delictivo de manera exponencial. La posibilidad de abrir cuentas bancarias remotamente, con procesos de verificación de identidad (KYC) automatizados y, por ende, vulnerables, y la consolidación de los sistemas de pago instantáneo 24/7 han conferido a las redes criminales una agilidad sin precedentes. Una transferencia que antes demoraba días y dejaba un claro rastro documental hoy puede ejecutarse y dispersarse por múltiples jurisdicciones en segundos, lo que complica enormemente la reacción de bancos y fuerzas de seguridad. Se suma a esto la notable capacidad de adaptación de los delincuentes, quienes explotan redes sociales y aplicaciones de mensajería para reclutar muleros de forma masiva y transfronteriza, apuntando a menudo a individuos sin historial delictivo que son engañados o cooptados.

La velocidad y el anonimato que los canales digitales proporcionan han convertido la ocultación de fondos no en una fase posterior y diferenciada, sino en una necesidad simultánea e inmediata al propio acto delictivo. Los fondos ilícitamente obtenidos deben moverse y estratificarse al instante para eludir su bloqueo y asegurar la impunidad de los perpetradores.

Es esta necesidad imperiosa de un blanqueo instantáneo la que ha creado una demanda voraz por un mecanismo específico: un vehículo intermediario que pueda recibir los fondos ilícitos y disociarlos de su origen a gran velocidad. Dicho eslabón crítico, el nexo indispensable que conecta el fraude con el blanqueo, es la “cuenta mula”. Se ha erigido en la pieza central de la arquitectura del crimen financiero contemporáneo, el punto de entrada mediante el cual los beneficios del fraude digital se inyectan en las cadenas de lavado de dinero.

En consecuencia, es imposible afrontar la prevención del fraude y las estafas digitales de un modo efectivo sin comprender y desarticular el fenómeno de las cuentas mula. Comprender esta simbiosis crítica resulta, en definitiva, el fundamento indispensable para poder desarrollar estrategias de defensa que no solo reaccionen ante las pérdidas, sino que ataquen la infraestructura que las posibilita.

2.2 CUENTAS MULA: DEFINICIÓN, OPERATIVA Y TIPOLOGÍAS DE MULEROS

En la parte central del blanqueo digital yace la "cuenta mula" un producto bancario ordinario que ha sido instrumentalizado para fines ilícitos. Es una cuenta bancaria, perteneciente a un titular legítimo, usada como vehículo intermediario para recibir y transferir fondos de origen delictivo. Su función primordial es actuar como cortafuegos en la cadena de transacciones, aislando el dinero de su origen fraudulento y, con ello, obstaculizando su trazabilidad. Por extensión, el "mulero" o "mula de dinero" es la persona física que, como titular o controlador de la cuenta, facilita estas operaciones, erigiéndose en el eslabón humano que permite a las redes criminales distanciarse del flujo financiero y minimizar su exposición legal.

La operativa de una cuenta mula constituye un engranaje clave en la fase de estratificación del blanqueo de capitales. El operativa habitual arranca con la recepción de fondos de una actividad ilícita. En vez de enviar el dinero a una cuenta propia, lo que les identificaría, los criminales lo desvían a la cuenta del mulero. Este, bajo instrucciones precisas, mueve el dinero de inmediato, sea transfiriéndolo a otras cuentas (frecuentemente otras mulas en cadena), retirándolo en efectivo o convirtiéndolo en activos líquidos y de difícil rastreo, como criptomonedas o tarjetas prepago. Cada salto añade una capa de opacidad; tras varias iteraciones, reconstruir el nexo entre el dinero y su origen se torna una tarea extremadamente compleja para los operadores.

La operativa de una cuenta mula revela una serie de patrones transaccionales atípicos que, a pesar de su sutileza, permiten su detección. Las cuentas mula tienden a mostrar una actividad marcada por la recepción súbita de fondos, a menudo de múltiples ordenantes desconocidos, seguida de una salida casi inmediata de la totalidad del dinero, lo que deja el saldo próximo a cero (fenómeno conocido como "nivelación de saldo"). Es común también la ausencia de una operativa personal ordinaria, como el pago de recibos o el ingreso de una nómina, lo cual evidencia su uso puramente instrumental.

La heterogeneidad de los perfiles de muleros impone una clasificación que supere la simple dicotomía entre culpables e inocentes. Organismos como Europol y el FBI, junto a la literatura académica, proponen una tipología que se basa en el nivel de consciencia e implicación del individuo:

Muleros ***Inconscientes*** o ***Engañados***: Conforman el grupo más numeroso y, a la vez, el más victimizado. Son personas sin antecedentes penales que ignoran su participación en una actividad ilícita. Suelen ser captadas mediante elaborados esquemas de ingeniería social, como falsas ofertas de trabajo, estafas románticas o supuestas oportunidades de inversión. Actúan convencidos de que realizan una tarea legítima. Su perfil "limpio" y su desconocimiento genuino los convierten en el vehículo idóneo para la primera fase del blanqueo, al ser menos propensos a activar las alertas bancarias.

Muleros ***Conscientes*** o ***Negligentes***: Este perfil habita una zona gris. Se trata de individuos que, sin conocer los detalles de la trama, sospechan de la irregularidad de las operaciones, pero optan por ignorar las señales de alarma, movidos por el beneficio económico o la indiferencia. Operan con una suerte de "ceguera voluntaria", justificando su participación con excusas. Legalmente, su conducta puede calificarse de cooperación necesaria o dolo eventual, pues asumen y aceptan el riesgo de estar colaborando con un delito.

Muleros ***Cómplices***: En este nivel se hallan los colaboradores plenamente conscientes e intencionales. Saben de la ilicitud de los fondos y participan activamente en la red de blanqueo a cambio de un beneficio directo o por lealtad a la organización. Con frecuencia gestionan múltiples cuentas, coordinan a otras mulas de nivel inferior incluso participan en el reclutamiento. Representan un eslabón más profesionalizado dentro de la red criminal.

Muleros ***Coaccionados***: Cabe añadir una categoría más: aquellos individuos forzados a participar bajo amenaza, deuda o explotación. Aquí se engloban víctimas de trata de personas, inmigrantes en situación irregular o personas extorsionadas. En tales casos, el individuo actúa sin libre albedrío, siendo doblemente víctima: de la red criminal y, potencialmente, del sistema judicial si su situación de coacción no es debidamente considerada.

Resulta fundamental entender que, con independencia del grado de consciencia, participar en estos esquemas acarrea graves consecuencias legales. Incluso el mulero más incauto puede afrontar cargos penales por blanqueo o encubrimiento, junto a la obligación de restituir los fondos y su inclusión en listas de exclusión financiera de por vida. Esta delgada línea entre víctima y perpetrador es uno de los mayores retos para entidades

financieras y autoridades, exigiendo no solo herramientas de detección avanzadas, sino también un enfoque matizado en la investigación y el enjuiciamiento.

2.3 CICLO DE VIDA DEL MULERO: FASES, PATRONES Y VULNERABILIDADES

La actividad de una cuenta mula no es un suceso aislado; forma parte de un proceso estructurado que puede analizarse como un "ciclo de vida". Dicho ciclo, que abarca desde la captación del individuo hasta su descarte por la red, se compone de tres fases: reclutamiento (onboarding), explotación activa (operativa) y desconexión (detección/huida). Cada etapa despliega patrones de comportamiento, tácticas y vulnerabilidades específicas, cuyo análisis resulta clave para el diseño de estrategias de prevención eficaces.

El reclutamiento, la primera fase, es crucial para la organización criminal, pues de ella depende su suministro constante de cuentas "limpias". Los métodos de captación han migrado masivamente al entorno digital, donde los reclutadores se apalancan en la escala y el anonimato de internet. La táctica más habitual son las ofertas de empleo fraudulentas, diseminadas en portales de trabajo o redes sociales, que prometen puestos de "trabajo desde casa" con altos ingresos y mínimo esfuerzo, como "agente de pagos". Tales ofertas suelen atraer a perfiles financieramente vulnerables (jóvenes, desempleados), más susceptibles a la promesa de dinero fácil.

Otra modalidad extendida son las estafas que apelan a vulnerabilidades emocionales. La captación también se da mediante mensajes directos en redes sociales que ofrecen comisiones por "prestar" la cuenta bancaria. Aunque menos frecuente, la captación física persiste en entornos como campus universitarios o centros de acogida, donde se puede ejercer una presión más directa. Los perfiles más buscados son generalmente jóvenes (menores de 35 años), con escasa educación financiera y sin antecedentes, cuyo historial limpio dificulta la detección inicial. Las vulnerabilidades explotadas aquí son, por tanto, principalmente humanas: precariedad económica, ingenuidad, soledad o miedo.

Una vez finalizado su proceso de onboarding de forma exitosa en la entidad bancaria el mulero, da comienzo la fase operativa, el núcleo del esquema. La cuenta empieza a recibir fondos de origen delictivo, redistribuidos de inmediato siguiendo instrucciones. La dinámica es casi siempre idéntica: a una entrada de dinero le sigue, en minutos u horas, una o varias salidas que agotan el saldo. Los métodos de salida varían: transferencias a otras mulas, conversión a criptoactivos, compra de tarjetas regalo o retiradas en efectivo. Durante esta fase, la cuenta exhibe patrones anómalos, inmediatez de los flujos, nivelación del saldo a cero, actividad internacional injustificada, que la distinguen de un cliente legítimo. Aquí, las vulnerabilidades explotadas son tanto tecnológicas como de procedimiento. La velocidad de los pagos instantáneos 24/7 permite que el dinero cruce varias jurisdicciones antes de que los sistemas de prevención puedan reaccionar. Los delincuentes además se benefician de la limitada comunicación entre entidades y países, una fragmentación que impide una visión completa de la red de mulas.

Finalmente, el ciclo concluye con la fase de desconexión. Cuando la cuenta ha sido utilizada o si emerge un riesgo de investigación, la red criminal "descarta" al mulero y corta cualquier vínculo. Los reclutadores cesan todo contacto, abandonándolo a su suerte. A menudo es en este punto cuando la persona engañada descubre la estafa, usualmente porque el banco bloquea su cuenta o le notifica una investigación. Para entonces, los fondos ya son irrecuperables. El mulero queda expuesto a todas las consecuencias: responsabilidad penal por blanqueo, la obligación de devolver el dinero defraudado (lo que puede generar una deuda perpetua, dado que su comisión es ínfima) y la inclusión en listas de exclusión financiera. La vulnerabilidad que los criminales explotan aquí es la lentitud de las investigaciones transfronterizas, confiando en que para cuando las autoridades rastreen el dinero, la pista se habrá enfriado y el mulero será el único eslabón identificable. Comprender este ciclo vital no solo ayuda a entender la mecánica delictiva, sino que revela puntos de intervención estratégicos, desde campañas de concienciación para frustrar el reclutamiento hasta la mejora de la cooperación internacional.

2.4 MARCO REGULATORIO Y MECANISMOS INSTITUCIONALES DE PREVENCIÓN

La creciente amenaza de las cuentas mula ha catalizado una notable evolución en los marcos regulatorios y mecanismos de control, tanto a escala nacional como internacional. La estrategia de respuesta se articula sobre la armonización legislativa, el refuerzo de la supervisión y el fomento de una cooperación público-privada más sólida para sellar las brechas que las redes criminales aprovechan.

A nivel global, las Recomendaciones del Grupo de Acción Financiera Internacional (FATF/GAFI) constituyen el estándar de referencia en la lucha contra el blanqueo de capitales y la financiación del terrorismo (AML/CFT) [3]. Aunque sus directrices carecen de vinculación jurídica, su influencia política es inmensa y sirve de cimiento para las legislaciones nacionales. El GAFI ha puesto un énfasis creciente en los riesgos del entorno digital, con guías sobre identificación no presencial de clientes, regulación de activos virtuales (AV) [4] y la necesidad de un enfoque basado en el riesgo adaptado a las nuevas tecnologías. Su insistencia en una acción global más enérgica contra los riesgos de los AV responde directamente a su uso ascendente en las etapas finales del blanqueo vía mulas.

En el contexto europeo, la Quinta Directiva Anti-Blanqueo (AMLD5), vigente desde 2020, marcó un hito al extender la regulación a plataformas de intercambio de criptomonedas y proveedores de monederos de custodia [5]. Esta medida les impuso las mismas obligaciones de diligencia debida (KYC) e informe de operaciones sospechosas que, a los bancos, buscando así reducir el anonimato que facilitaba la conversión de fondos ilícitos. La AMLD5 también reforzó la transparencia sobre la titularidad real de las empresas y limitó el uso de tarjetas prepago-anónimas. Por su parte, la Sexta Directiva (AMLD6) armonizó la definición de los delitos de blanqueo, incluyendo figuras como la complicidad y la negligencia, lo que facilita el encuadre penal de los muleros, incluso si estos alegan desconocimiento.

Consciente de que la fragmentación regulatoria seguía siendo una debilidad, la Unión Europea adoptó en 2024 un ambicioso paquete legislativo para crear la nueva Autoridad de Lucha contra el Blanqueo de Capitales (AMLA). Con sede en Frankfurt y operativa desde 2025, AMLA ostentará competencias de supervisión directa sobre las entidades de

mayor riesgo y coordinará a las Unidades de Inteligencia Financiera (UIF) nacionales para unificar criterios y agilizar el intercambio de información. Tal centralización responde directamente a la naturaleza transnacional de las redes de mulas. Por otro lado, la Directiva de Servicios de Pago (PSD2) ha tenido un impacto ambivalente. Si bien la Autenticación Reforzada del Cliente (SCA) ha elevado la seguridad y reducido el fraude técnico, también ha desplazado las tácticas criminales hacia la ingeniería social, donde se engaña a la víctima para que autorice ella misma el pago (fraude de pago autorizado o APP fraud). En estos esquemas, las cuentas mula son más cruciales que nunca para dispersar los fondos con celeridad.

A nivel nacional, España ha transpuesto estas directivas mediante la Ley 10/2010 de Prevención del Blanqueo de Capitales y de la Financiación del Terrorismo [6]. La ley impone a las entidades financieras estrictas obligaciones de diligencia debida y de comunicación de cualquier operativa sospechosa al SEPBLAC, la UIF española. En una iniciativa pionera, el SEPBLAC implementó en 2024 un nuevo procedimiento de comunicación agregada, que obliga a las entidades a reportar mensualmente y de forma estructurada todas las cuentas identificadas como posibles mulas [7]. Este sistema permite al SEPBLAC cruzar datos de todo el sector para identificar patrones y desvelar redes completas que operan de forma fragmentada.

Estos mecanismos se complementan con la cooperación interinstitucional y la sensibilización pública. Operaciones coordinadas por Europol como la Acción Europea contra las Cuentas Mula (EMMA) reúnen anualmente a policías y bancos de decenas de países, con miles de identificaciones y arrestos como resultado. A su vez, campañas de concienciación como #DontBeAMule, impulsadas por agencias como Europol e Interpol, buscan educar a la población sobre los riesgos de prestar una cuenta bancaria para así reducir la "cantera" de muleros [8]. Pese a este sólido andamiaje, persisten desafíos importantes, como el equilibrio entre la necesidad de compartir datos y el derecho a la privacidad (RGPD), y la velocidad con la que las redes criminales se adaptan a las nuevas tecnologías, un ritmo que a menudo desborda la respuesta regulatoria.

CAPÍTULO 3. LIMITACIONES DE LOS ENFOQUES ACTUALES DE DETECCIÓN

Para responder a la creciente sofisticación del fraude, los sistemas de detección han transitado desde arquitecturas basicas, basadas en reglas estáticas, hasta ecosistemas complejos que integran algoritmos de Machine Learning (ML) y Deep Learning (DL). Hoy, modelos como Random Forest, XGBoost o las redes neuronales constituyen el estándar tecnológico para la detección de fraude bancario, demostrando una capacidad superior para identificar patrones anómalos en volúmenes masivos de datos transaccionales. Su implementación, ciertamente, ha posibilitado una defensa más dinámica y precisa que la de sus predecesores.

A pesar de su potencia computacional, este paradigma exhibe signos de agotamiento estratégico. La agilidad con que las redes criminales adaptan sus tácticas, sobre todo en la orquestación de esquemas de cuentas mula, expone debilidades de los modelos de detección contemporáneos. Dichos sistemas, aun siendo técnicamente avanzados, operan a menudo sobre cimientos de datos imperfectos y con una perspectiva inherentemente reducida del fenómeno delictivo. Identificar una cuenta mula no es solo un problema de clasificación de transacciones; es la identificación de un actor en un ciclo de vida criminal y, con frecuencia, de un nodo dentro de una red coordinada. Justo en esta dimensión contextual y relacional los enfoques actuales revelan sus carencias más profundas.

Este capítulo emprende un análisis crítico de tales debilidades. Trasciende la simple evaluación comparativa de algoritmos para hurgar en las limitaciones conceptuales que lastran la eficacia real en la prevención del fraude.

Primero, exploraremos las deficiencias que nacen en la misma génesis del proceso: el tratamiento y la preparación de los datos. Seguidamente, se evaluarán con sentido crítico los motores de detección más extendidos, desde los sistemas de reglas hasta los modelos de ML, para desvelar sus fallos inherentes frente al desafío específico de las cuentas mula.

Por último, abordaremos el punto ciego sistémico que define el paradigma actual: una visión fragmentada y reactiva que, al enfocarse en eventos aislados, resulta incapaz de desarticular las estructuras criminales organizadas. El análisis no busca solo diagnosticar las insuficiencias del presente, sino erigir los pilares que justifican un cambio hacia un modelo holístico, proactivo y basado en redes, objeto de los capítulos posteriores.

3.1. DEBILIDADES EN EL ORIGEN: EL DATO COMO FUNDAMENTO COMPROMETIDO

Cualquier modelo predictivo depende, en su robustez y fiabilidad, de la calidad, completitud y representatividad de los datos que lo alimentan. Dentro del ámbito de la detección de fraude bancario, la fase de recopilación, tratamiento y preparación de datos, lejos de ser un trámite meramente técnico, introduce sesgos y limitaciones estructurales. Estos condicionan de manera irreversible el alcance y la eficacia de los sistemas de detección, constituyendo así la primera y más crítica barrera para una prevención verdaderamente efectiva.

Tales deficiencias en la calidad y estructura del dato no solo socavan la fiabilidad de los modelos predictivos, sino que, de forma crítica, lastran su capacidad de adaptación frente a un entorno delictivo dinámico.

Con el fin de ilustrar los puntos de fricción en este proceso, el siguiente esquema que representa la *Figura. 4* desglosa el flujo completo del tratamiento de datos en los sistemas de detección, abarcando desde la recolección inicial hasta la evaluación final del modelo.

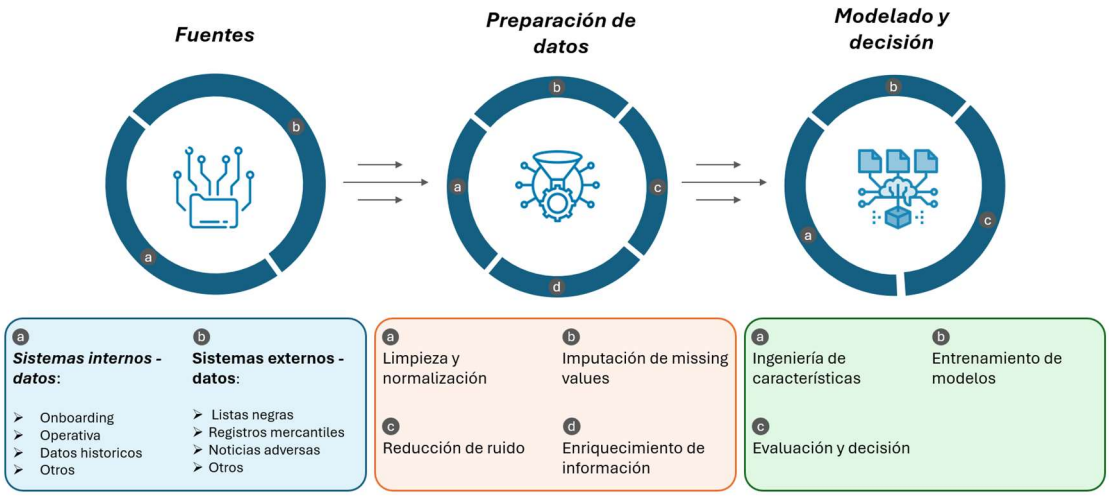

Figura 4. Fases de procesamiento de datos en la detección de fraude.

3.1.1. Falta de integración de datos y la visión fragmentada del cliente

El primer gran obstáculo yace en la arquitectura tradicional de los sistemas de información bancarios. La información del cliente no conforma un todo coherente y unificado; al contrario, se encuentra dispersa en múltiples repositorios departamentales, o "silos".

La gran mayoría de modelos de detección de fraude no opera sobre una visión 360 grados, sino sobre un subconjunto de datos limitados.

Tal fragmentación fuerza al modelo a decidir con información parcial. Detecta una transferencia anómala, sí, pero es incapaz de contextualizarla con el comportamiento no transaccional. La ausencia de una plataforma de datos unificada que integre y correlacione en tiempo real estas fuentes diversas supone una limitación estructural que empobrece la capacidad predictiva. El resultado es un modelo que identifica síntomas (la transacción sospechosa), pero no la causa (una cuenta instrumentalizada desde su origen). Superar estos silos no es solo un reto técnico de integración; es un requisito estratégico para construir el perfil de riesgo completo y dinámico del cliente, pilar del modelo de ciclo de vida que aquí se propone.

La fragmentación de la información representa un desafío medular para los sistemas antifraude. Lejos de ser un problema meramente técnico, la desconexión entre las distintas fuentes de datos genera cuellos de botella analíticos que merman la visibilidad tanto operativa como estratégica de las entidades.

3.1.2. El Desafío Crónico del Desbalance de Clases

En el inmenso universo de transacciones bancarias diarias, las operaciones fraudulentas y, en particular, las que involucran cuentas mula, constituyen eventos raros. Como reflejan numerosos estudios, su porcentaje sobre el total es ínfimo, representando a menudo menos del 0.2% del volumen [9]. Esta profunda asimetría estadística, conocida como "desbalance de clases", se erige como uno de los mayores desafíos técnicos y conceptuales al construir modelos de detección.

Los algoritmos de Machine Learning, por diseño, buscan minimizar el error global durante su entrenamiento. Ante un conjunto de datos tan desbalanceado, un modelo puede lograr una precisión muy alta, superior al 99.8%, con la simple estrategia de clasificar todo como legítimo. Aunque su "accuracy" numérica sea impecable, su utilidad práctica contra el fraude es nula. El modelo se vuelve un experto en la normalidad, pero muestra una ceguera notoria ante la minoritaria pero crítica clase de fraude.

Para mitigar este efecto, la industria ha recurrido a técnicas de balanceo. Los métodos de "submuestreo" (under-sampling), como bien se observa en la *Figura 5*., eliminan al azar instancias de la clase mayoritaria (legítimas) para equilibrar la proporción. Este enfoque, sin embargo, acarrea un riesgo considerable de perder información valiosa, al descartar patrones de normalidad que podrían ser clave para distinguir comportamientos limítrofes.

Por otro lado, las técnicas de "sobremuestreo" (over-sampling) aumentan el número de muestras de la clase minoritaria, también representado en la *Figura 5*. El método más conocido es SMOTE (Synthetic Minority Over-sampling Technique), que no solo duplica las muestras de fraude, sino que genera instancias sintéticas interpolando entre vecinos cercanos. Si bien esta técnica puede mejorar métricas como el "recall", sus inconvenientes son sustanciales. Generar datos sintéticos puede introducir ruido y crear patrones artificiales ajenos a la realidad del fraude, llevando al modelo a aprender artefactos en

lugar de señales de riesgo genuinas. Esto puede derivar en un "sobreajuste" (overfitting), donde el modelo rinde excepcionalmente con los datos de entrenamiento sintéticos, pero fracasa al generalizar a datos reales. Además, SMOTE puede crear solapamiento entre clases, haciendo la frontera de decisión más difusa e incrementando, en ciertos casos, la tasa de falsos positivos.

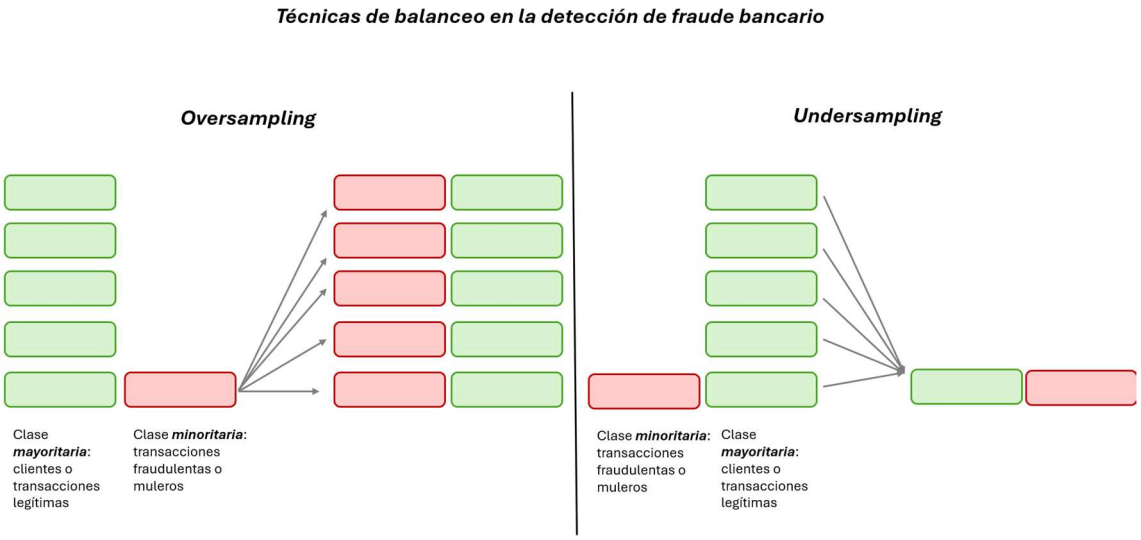

Figura 5. Técnicas de balanceo en la detección de fraude bancario.

Una alternativa más refinada es el uso de pesos de clase (class weights) en algoritmos como Logistic Regression o XGBoost, que penaliza más los errores en la clase minoritaria sin alterar los datos. Aun así, el problema de fondo subsiste: la escasez de ejemplos reales limita la capacidad del modelo para aprender la diversidad y complejidad de las tácticas criminales. Este desbalance no es un mero problema técnico para resolver con un algoritmo; es una limitación conceptual que clama por un cambio de enfoque. En lugar de depender solo de la detección de anomalías en datos escasos, se debe enriquecer el análisis con información contextual y relacional, como se propondrá.

3.1.3. Sesgos Reactivos en la Ingeniería de Características y el Concept Drift

La "ingeniería de características" (feature engineering) es el proceso por el cual los científicos de datos construyen las variables que un modelo de ML usará para aprender y

predecir. En la detección de fraude, estas suelen ser indicadores cuantitativos diseñados para capturar patrones de riesgo ya conocidos: la frecuencia de las transacciones, la hora, el importe, la novedad de la contraparte o la rapidez con que los fondos salen de la cuenta.

El problema inherente a este enfoque es su naturaleza "reactiva". Las características se diseñan a partir del análisis de fraudes que ya han ocurrido y sido identificados. Si se observa que las mulas vacían sus cuentas con celeridad, se crea una característica como `tiempo_entre_entrada_y_salida`. El modelo aprende a reconocer este patrón, volviéndose eficaz para detectar fraudes similares a los del pasado.

No obstante, las redes criminales son adversarios inteligentes, adaptativos. En cuanto una táctica es detectada y neutralizada de forma sistemática, la abandonan para desarrollar nuevas estrategias que eludan las defensas. Este fenómeno, conocido en el campo del ML como "concept drift", describe cómo la relación estadística entre las características y la variable objetivo (fraude) muta con el tiempo. Los patrones que ayer eran altamente predictivos de fraude mañana pueden volverse irrelevantes o generar falsas alarmas.

Los modelos entrenados con datos históricos se degradan, por tanto, de forma inevitable. Inicialmente una cuenta mula empieza su operativa con pequeñas transacciones legítimas para construir un historial de normalidad antes de la transacción fraudolenta, invalidando las características basadas en la actividad súbita. O pueden coordinar una red para dispersar fondos en importes mínimos (smurfing), eludiendo los umbrales de alerta basados en grandes volúmenes.

Esta dependencia de una ingeniería de características anclada en el pasado condena a los sistemas a una perpetua carrera por detrás del defraudador. Para romper este ciclo, es imperativo que los modelos no solo se centren en variables transaccionales predefinidas, sino que sean capaces de analizar el comportamiento del cliente a lo largo de su "ciclo de vida" y las "relaciones" que establece, identificando así anomalías contextuales que pueden incluso preceder a la actividad fraudulenta.

3.1.4. Las Restricciones de la Privacidad y la Colaboración de Datos

Finalmente, una limitación cada vez más definitoria es el marco regulatorio sobre la privacidad de datos, con el Reglamento General de Protección de Datos (GDPR) en Europa como su máximo exponente. Si bien estas normativas son cruciales para proteger los derechos ciudadanos, imponen restricciones significativas a cómo las entidades financieras pueden recopilar, procesar y, sobre todo, compartir información.

El problema de las redes de mulas es, por definición, transbancario y transfronterizo. Una misma organización criminal puede operar con cuentas en docenas de bancos distintos por toda Europa. Su detección efectiva exigiría una colaboración y un cruce de datos masivo entre entidades

Cada banco posee una visión aislada de la red. Puede identificar algunas mulas en su cartera, pero ignora que están conectadas a otras tantas de otras entidades. Esta falta de visibilidad global es el principal activo de las redes criminales, que explotan la fragmentación del ecosistema para diluir sus operaciones y evadir la detección.

Aunque existen iniciativas para fomentar el intercambio de información bajo marcos seguros, como las impulsadas por Europol o el sistema de comunicación agregada del SEPBLAC en España, sus soluciones tienden a ser lentas o se basan en información agregada, perdiendo el detalle granular necesario para un análisis de red en tiempo real. Tecnologías emergentes como el "aprendizaje federado" (federated learning) prometen una salida a este dilema, al permitir entrenar un modelo global sobre datos descentralizados sin que los datos brutos salgan de cada entidad. Con todo, su implementación a gran escala en el sector financiero es aún incipiente y compleja.

Esta limitación regulatoria refuerza la necesidad de que cada entidad maximice el valor de los datos que "sí" controla. Si no es factible ver la red interbancaria completa, se vuelve aún más crucial poder construir y analizar la red de interacciones internamente al propio banco.

3.2. LAS LIMITACIONES DE LOS MOTORES DE DETECCIÓN

Una vez los datos son recopilados y procesados, con todas las limitaciones inherentes que ello supone, entran en juego los motores de detección. Estos son los modelos encargados de discernir la señal del ruido y emitir un juicio de riesgo. Históricamente, dichos motores han seguido una clara trayectoria evolutiva, desde sistemas deterministas basados en reglas hasta complejos modelos de aprendizaje automático. No obstante, a pesar de la creciente sofisticación, cada paradigma arrastra debilidades conceptuales que merman su eficacia en el dinámico y adversarial contexto de las cuentas mula.

3.2.1. LA RIGIDEZ Y OBSOLESCENCIA DE LOS MOTORES DE REGLAS

Los sistemas basados en reglas y umbrales fueron la primera generación de defensas automatizadas. Su lógica es intuitiva, determinista y, sobre todo, transparente. Se componen de una serie de sentencias condicionales (`IF-THEN-ELSE`) definidas por expertos. Por ejemplo: `Si una cuenta con <30 días de antigüedad recibe >3.000€ de un ordenante desconocido y realiza múltiples salidas en las siguientes 2 horas, entonces generar alerta`.

La principal virtud de estos sistemas es su “interpretabilidad”. Un analista comprende de inmediato por qué se generó una alerta, facilitando la investigación y la justificación de las acciones. Además, su implementación es relativamente simple y su coste computacional, bajo.

Sus limitaciones, sin embargo, son varias y diferentes. La primera de estas es su “rigidez inherente”. Las reglas son estáticas, binarias; una transacción cumple o no la condición, sin espacio para la ambigüedad o patrones sutiles. Las redes criminales, como adversarios racionales, dedican un esfuerzo considerable a entender la lógica de estos sistemas para poder eludirla. Aprenden los umbrales de importe, frecuencia o tiempo y ajustan su operativa para pasar desapercibidos. Si la alerta salta a los 3.000€, fraccionan los fondos en transferencias de 2.900€. Si se activa por actividad en cuentas nuevas, usan cuentas "durmientes" o las "calientan" previamente.

Conforme los analistas descubren tácticas y añaden reglas, el sistema se vuelve más complejo y difícil de mantener. Un exceso de reglas puede generar interacciones inesperadas. Además, estos sistemas son conocidos por su alta tasa de falsos positivos. Reglas muy genéricas pueden bloquear un gran número de transacciones legítimas, provocando una fricción inaceptable para el cliente y erosionando la confianza. Un cliente al que se le bloquea una transferencia urgente legítima percibe al banco no como un protector, sino como un obstáculo. Este volumen de falsas alarmas satura a su vez a los equipos de investigación, que dedican tiempo a revisar alertas de bajo riesgo, pudiendo ignorar las verdaderas amenazas. En esencia, los motores de reglas son una herramienta necesaria, pero del todo insuficiente por sí sola.

Las reglas son una herramienta de gran utilidad para el análisis post-mortem. Al recibir una denuncia sobre cuentas mula, la revisión de sus activaciones permite entender el flujo del fraude y reconstruir su "storytelling". No obstante, su eficacia para la detección proactiva es muy limitada. Este método, por sí solo, difícilmente permite anticiparse a la comisión del fraude.

3.2.2. MACHINE LEARNING

La irrupción del Machine Learning (ML) supuso el avance más significativo en la lucha contra el fraude financiero en las últimas dos décadas. Este salto cualitativo desplazó el paradigma desde una lógica determinista, fácilmente eludible por adversarios metódicos, hacia un enfoque probabilístico y adaptativo, capaz de identificar patrones complejos y no lineales en volúmenes masivos de datos. La promesa del ML consistía en superar la rigidez de las reglas “if-then-else” para descubrir correlaciones sutiles y comportamientos anómalos que un analista, o un sistema de reglas, sería incapaz de detectar.

Modelos como “Random Forest”, “Gradient Boosting” (XGBoost), Redes Neuronales o “Isolation Forest” se han convertido en el estándar de la industria, demostrando una capacidad superior para procesar miles de características de forma simultánea. En el contexto de las cuentas mula, esto se traduce en la habilidad de analizar no solo el importe de una transacción, sino también su frecuencia, su relación con la actividad histórica del

cliente, la reputación de la contraparte o el dispositivo desde el que se ordena, generando así una puntuación de riesgo multidimensional y dinámica.

El aprendizaje automático ofrece un abanico de metodologías para resolver problemas complejos, cuyo enfoque se adapta tanto al tipo y volumen de los datos disponibles como a la naturaleza del objetivo analítico.

La *Figura 6.* a continuación presenta una clasificación de las ramas fundamentales del Machine Learning, y evidenciadas en rojos están las técnicas que son mas utilizadas en la detección de fraude en banca.

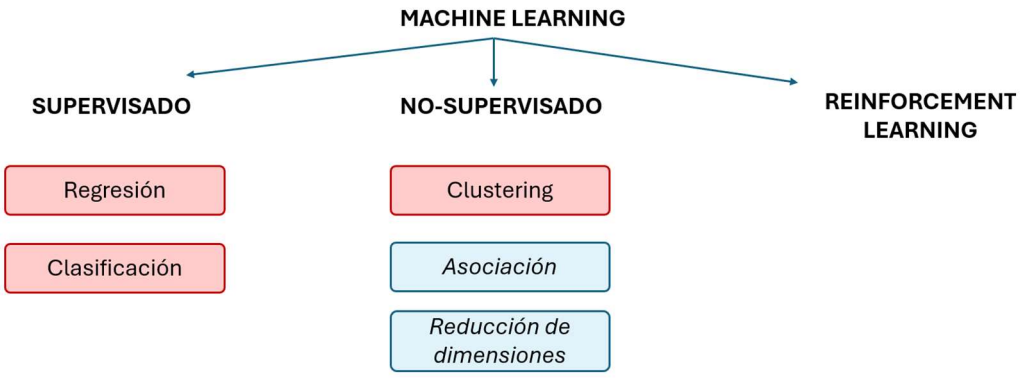

Figura 6. Clasificación general de métodos de Machine Learning.

Sin embargo, su innegable potencia no ha estado exenta de contrapartidas. La implementación de modelos de ML ha desvelado un nuevo espectro de desafíos, tan o más complejos que los que pretendía resolver. Estos sistemas, aunque técnicamente avanzados, no son una panacea. Su eficacia está intrínsecamente ligada a la calidad de los datos con los que se entrenan, su capacidad para adaptarse a un entorno adversarial en constante cambio y, de forma crítica, su interpretabilidad ante reguladores y equipos operativos.

El problema fundamental radica en que muchos de estos modelos operan bajo una "visión de túnel transaccional", con el objetivo único de clasificar un evento discreto (una transacción, una sesión) como "fraudulento" o "legítimo". Esta perspectiva, aunque útil,

es inherentemente reduccionista. Ignora que una cuenta mula no es un evento aislado, sino la manifestación de un proceso criminal con un ciclo de vida estructurado y que a menudo funciona como un nodo dentro de una red coordinada.

Este apartado se adentra en un análisis crítico de los motores de detección basados en Machine Learning, exponiendo sus fallos conceptuales y limitaciones estructurales en el contexto específico de las redes de mulas. Se evaluarán las técnicas más extendidas a la luz de la evidencia empírica y se explorarán los desafíos sistémicos, desde el desbalance de clases y la opacidad de la "caja negra" hasta las restricciones éticas y de privacidad, que definen el estado del arte. El objetivo no es desacreditar el ML, sino realizar un diagnóstico riguroso de sus carencias para justificar la necesidad de evolucionar hacia el paradigma holístico y relacional que se propondrá en los capítulos posteriores.

3.2.2.1. Modelos avanzados para la Detección de Mulas: Un Análisis Detallado

La industria financiera ha adoptado un diverso conjunto de algoritmos de ML, cuya elección depende de factores como la naturaleza de los datos, los recursos computacionales, los requisitos de interpretabilidad y la madurez del equipo de ciencia de datos. A continuación, se analizan las familias de modelos más relevantes.

Modelos Basados en Ensambles de Árboles: Random Forest y Gradient Boosting (XGBoost)

Los métodos de ensamble, que combinan múltiples modelos simples para obtener una predicción más robusta, son los más populares y efectivos en la detección de fraude.

El algoritmo *Random Forest (RF)* construye un gran número de árboles de decisión durante el entrenamiento. Para cada árbol, utiliza una muestra aleatoria de los datos y de las características, y la predicción final agrega los resultados de todos los árboles. En la detección de mulas, un RF puede analizar simultáneamente cientos de “features” como 'importe_transacción' o 'antigüedad_cuenta'. Su robustez frente al ruido lo hace ideal para datos financieros complejos y diversos estudios demuestran su alto rendimiento. A pesar de ello, su lógica interna puede ser opaca, contribuyendo al problema de la "caja negra",

como se puede observar en la *Figura.7*, y su eficacia se degrada frente a nuevas tipologías de fraude no vistas en el entrenamiento.

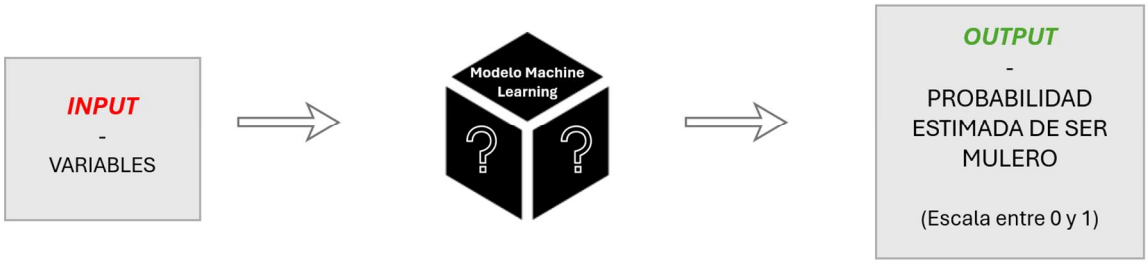

Figura 7. Esquema de "caja negra".

Por su parte, los modelos de *Gradient Boosting Machines (GBM*), y en particular *XGBoost*, construyen los árboles de forma secuencial, de modo que cada nuevo árbol corrige los errores de los anteriores. XGBoost es una implementación optimizada que a menudo supera a Random Forest en los benchmarks de la industria. Su eficiencia computacional y sus técnicas de regularización para prevenir el sobreajuste lo hacen muy valioso. Puede ser especialmente efectivo para identificar patrones sutiles al dar más peso a las transacciones anómalas mal clasificadas. Informes del sector destacan que XGBoost es el modelo elegido por algunas entidades para sus sistemas de detección de fraude en tiempo real debido a su baja latencia. No obstante, es más sensible a los hiperparámetros que RF y su interpretabilidad es igualmente un desafío.

Modelos de Detección de Anomalías (Aprendizaje No Supervisado): Isolation Forest y Autoencoders

Dado que el fraude es un evento raro, el aprendizaje no supervisado ofrece un enfoque complementario. En lugar de usar etiquetas, estos modelos aprenden qué es un comportamiento "normal" y señalan cualquier desviación significativa.

El algoritmo *Isolation Forest (IF)* parte de la premisa de que las anomalías son "pocas y diferentes", lo que las hace más fáciles de aislar. Su gran ventaja es la eficacia en la

detección de nuevas tipologías de fraude, lo que lo hace intrínsecamente más robusto contra el concept drift. Estudios relevantes han demostrado que, mientras el rendimiento de un Random Forest se desploma al introducir una nueva anomalía, el Isolation Forest mantiene un rendimiento estable. Esto lo convierte en una herramienta estratégica para descubrir amenazas emergentes, aunque puede tener dificultades en conjuntos de datos de muy alta dimensionalidad.

Un *Autoencoder (AE)* es una red neuronal entrenada para reconstruir su propia entrada. Al entrenarse sobre datos mayoritariamente legítimos, aprende a reconstruir bien las transacciones normales, pero falla al intentar reconstruir una anómala. La anomalía se detecta midiendo el "error de reconstrucción". Aunque son potentes para aprender patrones complejos y no lineales, comparativas académicas han mostrado un rendimiento inferior al de Isolation Forest. Además, su entrenamiento es computacionalmente más costoso y son modelos inherentemente opacos.

3.2.2.2. ANÁLISIS COMPARATIVO Y EVIDENCIA EMPÍRICA

La elección de un modelo de Machine Learning no es trivial, y la literatura académica ofrece una perspectiva valiosa sobre el rendimiento comparativo de distintas técnicas. Para fundamentar el análisis de las limitaciones inherentes a estos enfoques, es crucial examinar la evidencia empírica. En este sentido, la revisión sistemática de Soria, Loayza Abal y Segura Peña (2024) en el paper intitulado: “Machine Learning Models for Money Laundering Detection in Financial Institutions. A Systematic Literature Review” [10] constituye una fuente de excepcional valor, al sintetizar la investigación global sobre modelos de ML para la detección de blanqueo de capitales entre 2015 y 2023. El análisis de las tablas extraídas de dicho estudio permite trazar un mapa preciso del estado del arte.

La Tabla 1 que resume los artículos de revisión, revela dos tendencias clave. En primer lugar, la naturaleza global del problema es evidente, con contribuciones académicas de China, India, Estados Unidos y diversas naciones europeas, lo que refleja la dimensión transnacional de la amenaza y la consiguiente necesidad de una respuesta investigadora coordinada. En segundo lugar, la tabla muestra una extraordinaria diversidad de algoritmos bajo escrutinio, desde modelos clásicos como árboles de decisión y máquinas

de vectores de soporte (SVM), hasta enfoques avanzados como redes neuronales convolucionales (CCNN). Esta heterogeneidad denota una intensa actividad investigadora en busca de la arquitectura algorítmica más eficaz.

Tabla 1. Mapa descriptivo del panorama de la investigación actual [10]

Title	Reference	Country	Models ML
Una A time and frequency-based detection of suspicious activity to combat money laundering [9].	Ketenci, Utku Gorkem et atl (2021)	Turquía	Transaction Feature; Time Frequency and CRM Features
Intelligent anti-money laundering fraud control using a graph-	Naveed, Nasir et al (2022)	Pakistán	Decision Tree (DT), Conditional Inference Tree (CT), Random

based machine learning model for the financial domain [10]			Forest (RF); Neural Network (NN)
CCNN: CCNN: an artificial intelligence-based classifier for a credit card fraud detection system with an optimized cognitive learning model. [11]	Vetrivendan L. et al Mayo (2023)	Noida, India	proposed cognitive convolutional neural network (CCNN) classifier. Existing classifiers such as logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT) and support vector machine (SVM) were used to classify the proposed classifier. (SVM)
Dominant feature selection and machine learning-based hybrid approach to analyze Android ransomware. [12]	Gera, Tania et al. (2021)	Punjab, India	J48, Random Forest, LMT, Random Tree
Machine learning with belief rule-based expert systems to predict stock price movements. [13]	Emam Hossain et al. (2022)	Baltimore, EE.UU	Belief-Based Rule-Based Expert System (BRBES)
Application of the deep learning method in the TVP-VAR model under systematic financial risk monitoring and early warning. [14]	Huang, Anzhong et al. (2023)	China	Deep learning ; Early warning ; Information systems ; Supervision ; Information systems ; Systemic financial risk
Development of a predictive customer investment model using the conjoint learning technique. [15]	Kaewkiriya, Thongchai et al. (2022)	Thailand	Clustering, K-Nearest Neighbour Algorithm, Naïve Bayes Algorithm, Decision Tree Algorithm, Neural Network Algorithm,
Validating the impact of accounting disclosure on the stock market: a deep neural network approach. [16]	Eachempati, Prajwal (2023)	India	Deep neural networks with LSTM, Naive Bayes, Maximun Entropy, SVM, RNNR
Deep Learning criminal networks. [17]	Ribeiro, Haroldo V et al. (2023)	Brasil	Convolutional networks; GraphSAGE
DeLClustE: Protecting users against credit card transaction fraud through deep learning cluster ensemble. [18]	Aghware, Fidelis Obukohwo et al. (2023)	Agbor, Nigeria	DNN, PHMM, MNN, GANN ,DelCluste.
Detection of manipulators in cryptocurrency markets based on forecast anomalies. [19]	Akba, Firat et al. (2023)	Turquía	SARIMAX, ARIMA, LSTM, SVM
Research on the application of machine learning for watch list filtering in the fight against money laundering. [20]	Qutqut, H et al. (2023)	Jordania	SVM, DT Y NB (Decision Tree) Naïve Bayes, Support Vector Machine
Unbalanced classification of fraudulent bank transactions using machine learning. [21]	Ruchay, Alexey et al. (2020)	Federation de Rusia	algorithms TPOT y Random Forest
Transactional network analysis and identification of China's central bank digital currency	Li, Ziyu et al. (2020)	China	GCN, EvolveGCN, GAT GraphSAGE, ChebNet-GRU

money laundering behavior. [22]			
Multilayer perceptron artificial neural network-based model for credit card fraud detection. [23]	Kasasbeh, Bassam (2020)	Jordania	Best Network Model ,
Money laundering detection using machine learning and deep learning. [24]	Alotibi, Johrha (2019)	Arabia Saudita	NB, RF, KNN , DNN
Exploiting machine learning algorithms to detect financial crime based on customer behavior. [25]	Kumar, Sanjay et al. (2022)	Suiza	decision tree (DT), random forest (RF) and k-nearest neighbor (KNN).
Predictive financial fraud detection analytics using Azure and Spark ML.[26]	Purushu, Priyanka et al. (2018)	Estates Unidos	LR, DF , DJ , SVM
A flow-based approach for Trickbot banking Trojan detection. [27]	Gezer, Ali et al. (2019)	Turquía	random forest, multilayer perceptron's, minimal sequential optimization and Logistic Models
Money laundering risk assessment of bank accounts using naive bayes classification. [28]	Islam, MA et al (2020)	Bangladesh	Level Search Method (RLFM) in the context of Money and Laundering Residence in Naive Bayes.
Research on the application of machine learning for watch list filtering in anti-money laundering. [29]	Asha RB, et al (2021)	India	Support Vector Machine (SVM), k-nearest neighbor (KNN) and artificial neural network (ANN)
Machine learning approaches for the construction of the national anti-money laundering index.[30]	Zhang, GK, et al (2023)	China	LASSO regression and random forests
Credit card fraud detection using a new hybrid machine learning architecture. [31]	Malik, EF, et al (2022)	Malasia	hybrid machine Learning models.
Detection of money laundering and terrorist financing using neural networks and an anomaly indicator. [32]	Rocha-Salazar, JDJ et al (2021)	España	integrated model
Money laundering governance and income transfer: evidence from Australian financial institutions..[33]	Baban Eulaiwi et al (2024)	Australia	Asset Laundering Control with AI

Si bien la primera tabla informa sobre qué se investiga y dónde, la segunda es necesaria para evaluar la eficacia de dichos enfoques. La *Tabla. 2*, que presenta los resultados de rendimiento, ofrece una visión cuantitativa a través de métricas como el F1-Score y la Precisión del Modelo. El examen de estos datos permite extraer conclusiones más profundas, aunque no exentas de matices.

Tabla 2. Resultados de los artículos seleccionados por rendimiento [10]

Models used in money laundering	F1 SCORE	Model accuracy
Transaction Feature; Time Frequency and CRM Features.	59.37%; 72.19%	
Decision tree (DT), conditional inference tree (CT), random forest (RF); neural network (NN)	0.423 (DT) , 0.205 (CT) , 0.524 (RF), 0.414 (NN) . Acuracy 0.637 (DT) , 0.557 (CT) , 0.678 (RF), 0.693 (NN)	0.637 (DT), 0.557(CT), 0.678 (RF) , 0.693 (NN)
Proposed cognitive convolutional neural network (CCNN) classifier. Existing classifiers such as logistic regression (LR), K-nearest neighbor (KNN), decision tree (DT) and support vector machine (SVM) have been used. (SVM)	94% (LR), 93% (KNN , 93% (SVM), 90% (DR), 95.6% Cognitive CNN	Logistic Regression LR (94%), Knowledge nearest neighbour KNN (93%), Support Vector Machine SVM (93), Decision Tree Classifier DTC (90%) Cognitive CNN (CCNN) 95.6%

J48, Random Forest, LMT, Random Tree		J48 (0.97734), Random Forest (0.9863), LMT (0.9847), Random Tree (0.9837)
Belief-Based Rule-Based Expert System (BRBES)	BRB 93.50 ; RMSE 0.1233, R^2=48.06 AUC=0.984	
Deep learning ; Early warning ; Information systems ; Supervision ; Information systems ; Systemic financial risk		
Clustering, K-Nearest Neighbor Algorithm, Naïve Bayes Algorithm, Decision Tree Algorithm, Neural Network Algorithm,		
Deep Neural Networks with LSTM, Naive Bayes, Maximun Entropy, SVM, RNNR		Naive Bayes (0.64), SVM (0.67), RNAR(0.694), RNA Largo plazo (0.72)
Convolutional networks; GraphSAGE	0.88 ; 0.92 R^2 ajustada de 0,64 a 0,90	
DNN, PHMM, MNN, GANN ,DelClus te.		DNN(0.92) PHMM (0.89) MNN(0.91) , GANN (0.78) ,DelCluste (0.96)
SARIMAX, ARIMA, LSTM, SVM	ARIMA (62.5) , SARIMAX (64.5), SVM (60.00) LSTM (60)	ARIMA (63) , SARIMAX (60), SVM (83) LSTM (70)
SVM, DT Y NB (Decision Tree) Naïve Bayes, Support Vector Machine		SVM (0.815), NB (0.804), DT (0.782)
TPOT and Random Forest algorithms	RandomForestClassi fier (0.9610), TpotClassifier (0.9620)	RandomForesrClassifier(0.99 99), TpotClassifier (0.9999)
GCN, EvolveGCN, GAT GraphSAGE, ChebNet-GRU	GCN(0.960), EvolveGCN(0.961), GAT (0.962) GraphSAGE(0.968), ChebNet-GRU(0.971)	GCN(0.805), EvolveGCN(0.919), GAT (0.887) GraphSAGE(0.929), ChebNet-GRU(0.943)
Best Network Model ,	Best Network Model Acuracy (99.9505), RMSE(0.0218), F1(99.949, Specificity (79.710), AUC 0.8983	Fraud Measure F were 84,76%, 85,13% y 82,51% in one hidden layer, two hidden layers and three hidden layers,

NB, RF, KNN , DNN	NB(0.74) RF (0.99) KNN (0.97) DNN (0.98)	NB(0.99) RF (0.99) KNN (0.97) DNN (0.98)
decision tree (DT), random forest (RF) and k-nearest neighbor (KNN).	NB(0.90), DT (0.83) KNN(0.89) RF(0.90) SVM(0.90) LR(0.90)	NB(0.96), DT (0.83) KNN(0.96) RF(0.99) SVM (1) LR(0.99)
LR, DF , DJ , SVM	LR(1.000) DF (0.727) DJ (1.000) SVM (1.000)	LR(0.991) DF (0.995) DJ (0.997) SVM (0.993)
random forest, multilayer perceptron's, minimal sequential optimization and Logistic Models	random forest .0.939 multilayer perceptron's 0.667 minimum sequential optimization 0.997 Logistic Models . 0.998	random forest.0.999 multilayer perceptron's 0.997 minimum sequential optimization 0.995 Logistic Models . 0.996
Level Search Method (RLFM) in the context of Money and Laundering Residence in Naive Bayes	RLFM 0.94 Ingenuo Bayes 0.86	Naïve bayes 0.874
Model (SVM), k-nearest neighbor (Knn) artificial neural network (ANN)	SVM 94% KNN 92% ANN 90%	SVM 0.93 KNN 0.90 ANN 0.91
LASSO regression and random forests	Lasso 0.87 Random Forest 0.92	Lasso 0.976 Random Forest 0.987
hybrid machine learning models	Híbrido 0.86	H 0.987
integrated model	MI 0.87	MI 0.867
Deep Learning Artificial Intelligence Models	RNN 98.5%	Neural networks

A primera vista, los resultados parecen muy prometedores. Modelos como Random Forest o SVM alcanzan precisiones reportadas que a menudo rozan el 99%, y enfoques de Deep Learning como ChebNet-GRU también superan el 96%. Como concluyen los propios autores de la revisión, los modelos SVM y las Redes Neuronales tienden a presentar, en promedio, las tasas de precisión más elevadas. Esto confirma la capacidad teórica del ML para clasificar transacciones con un alto grado de exactitud.

No obstante, es imperativo interpretar estos valores con una perspectiva crítica. La tabla agrega resultados de estudios heterogéneos, cada uno con sus propios conjuntos de datos y metodologías. Por tanto, una comparación directa de las cifras, afirmar que el 99.99% de un Random Forest en un estudio es categóricamente superior al 95.6% de una CCNN en otro, sería metodológicamente incorrecta. El verdadero valor de la tabla

no reside en proclamar un "algoritmo ganador", sino en ilustrar el rango de eficacia que estas tecnologías pueden alcanzar, sirviendo como base para una discusión sobre sus limitaciones compartidas.

De hecho, esta evidencia refuerza la tesis central de este capítulo. Emerge una paradoja: los modelos que frecuentemente reportan la mayor precisión, como las redes neuronales profundas, son precisamente los más opacos. Su complejidad interna, que les permite capturar patrones sutiles, es la misma cualidad que los convierte en "cajas negras", enfrentándonos directamente al dilema de la interpretabilidad. Una precisión del 96%, si bien impresionante, resulta insuficiente si un analista no puede justificar por qué una cuenta ha sido intervenida.

En conjunto, ambas tablas dibujan el retrato de un campo de investigación vibrante y tecnológicamente sofisticado, con resultados notables en entornos controlados. Sin embargo, esta aparente madurez oculta debilidades sistémicas. La alta precisión reportada por modelos opacos subraya el desafío de la IA Explicable (XAI), mientras que la diversidad de enfoques refleja una búsqueda continua de soluciones a problemas persistentes como el concept drift y el desbalance de clases, desafíos que limitan la robustez de estos sistemas en el mundo real.

El análisis de esta sección revela una paradoja fundamental: los modelos de Machine Learning son las herramientas más potentes que poseemos para combatir el fraude, pero su paradigma de aplicación actual sufre de limitaciones conceptuales que merman su eficacia estratégica.

Hemos visto cómo, a pesar de su sofisticación, los modelos convencionales luchan con la escasez y el desbalance de los datos. operan como "cajas negras" que generan desconfianza, son inherentemente reactivos frente a adversarios adaptativos y están constreñidos por una visión fragmentada que les impide ver las redes criminales completas.

La raíz común de estas limitaciones es el enfoque predominante en la clasificación de eventos discretos. Al plantear el problema como "¿es esta transacción una mula?", se fuerza al modelo a operar con una visión de túnel, perdiendo el contexto temporal y relacional. Una cuenta mula no es una transacción, es una narrativa criminal que se despliega a través de un ciclo de vida y dentro de una red de conexiones.

Por tanto, para superar estas debilidades no basta con un algoritmo marginalmente mejor. Se requiere un cambio de paradigma. Es imperativo pasar de la clasificación de eventos a la identificación de comportamientos y trayectorias anómalas a lo largo del tiempo. Necesitamos modelos que no solo vean la transacción, sino que comprendan el ciclo de vida completo del cliente y analicen las relaciones que establece.

Este diagnóstico crítico justifica la necesidad de explorar el modelo holístico, proactivo y relacional que se detallará en los capítulos 4 y 5, un modelo que pone el ciclo de vida del cliente y el análisis de grafos en el epicentro de la estrategia de defensa.

3.3. UN PARADIGMA FRAGMENTADO Y REACTIVO

La confluencia de las debilidades en el tratamiento de datos y las limitaciones de los motores de detección genera un problema sistémico de mayor calado que define al paradigma actual de lucha contra el fraude: este es, en esencia, fragmentado y reactivo. Las defensas se erigen como un archipiélago de soluciones puntuales que responden a amenazas ya materializadas, en lugar de operar como un sistema de inteligencia proactivo y unificado, capaz de anticipar y desarticular las operaciones criminales en su totalidad.

3.3.1. REACTIVIDAD VS. PROACTIVIDAD: LA LUCHA DESPUÉS DEL DAÑO

La reactividad es la consecuencia lógica de la visión de túnel transaccional y la ingeniería de características basada en el pasado. La inmensa mayoría de sistemas de detección actuales entra en acción cuando el dinero ya ha sido sustraído y está en movimiento a través de la cuenta mula. En ese momento, el fraude ya se ha consumado y el blanqueo ha comenzado. La detección se produce en la fase de explotación activa del ciclo de vida del mulero.

Aunque bloquear una transacción o congelar una cuenta mula activa es una contención necesaria, representa una victoria pírrica. El daño financiero inicial ya está hecho y la recuperación de fondos suele ser una tarea ardua con pocas probabilidades de éxito, sobre todo si el dinero se dispersa por múltiples jurisdicciones o se convierte en criptoactivos. La entidad se ve abocada a un costoso proceso de investigación, gestión de reclamaciones y posible restitución, sumado al daño reputacional.

Un paradigma verdaderamente efectivo ha de ser proactivo. La proactividad implica desplazar el foco desde la explotación hacia las etapas más tempranas del ciclo criminal. El objetivo estratégico no es solo interceptar el dinero, sino impedir que la cuenta sea usada como instrumento delictivo.

3.3.2. La Fragmentación y la Ceguera ante las Redes Criminales

El problema de la fragmentación es, quizás, la vulnerabilidad sistémica más crítica, la que las organizaciones criminales explotan con mayor eficacia. El crimen organizado no se apoya en muleros individuales; su fuerza reside en orquestar redes de cuentas mula complejas, distribuidas y a menudo internacionales.

Un modelo de detección convencional, al analizar cada cuenta o transacción de forma independiente, es estructuralmente ciego a estas redes. Su unidad de análisis es el nodo (la cuenta), no las conexiones (las relaciones). Puede que una sola cuenta mula, con una transacción sospechosa, no acumule riesgo suficiente para generar una alerta prioritaria. Su actividad, aislada, podría parecer una anomalía de bajo impacto.

Sin embargo, si se amplía la perspectiva y se analiza el contexto relacional, la imagen cambia por completo. Esa misma cuenta podría estar conectada a otras cincuenta abiertas en un corto período, accedidas desde el mismo rango de IPs, que transfieren fondos a un mismo destino (p. ej., una plataforma de criptoactivos) y cuyos titulares comparten rasgos demográficos. La visión de esta estructura coordinada no deja duda: no son anomalías aisladas, sino una operación de blanqueo organizada a escala industrial.

Esta ceguera ante las redes es el principal punto ciego de la industria. Los modelos de ML y DL tradicionales, pese a su complejidad, carecen de la capacidad nativa para un análisis relacional. No están diseñados para interpretar grafos de conexiones, identificar comunidades de comportamiento anómalo o medir la centralidad de un nodo en una red. Son excelentes clasificadores de eventos individuales, pero tienen el límite en la detección de conspiraciones.

Las organizaciones criminales son plenamente conscientes de esta debilidad y la explotan sistemáticamente, dispersando sus operaciones entre múltiples cuentas y entidades para

que cada fragmento, por sí solo, parezca insignificante. Saben que los bancos rara vez tienen la capacidad de unir los puntos para ver el panorama completo.

Superar esta fragmentación es el desafío más urgente y el área donde una nueva estrategia puede aportar el máximo valor. Requiere un cambio fundamental desde la clasificación de eventos al análisis de grafos y redes. Un enfoque que modele las relaciones entre clientes, dispositivos, cuentas y transacciones puede desvelar las estructuras ocultas del crimen organizado, invisibles para los modelos actuales. La detección ya no se centraría en "¿es esta cuenta una mula?", sino en "¿forma esta cuenta parte de una red de blanqueo?".

En resumen, los enfoques actuales para detectar cuentas mula, a pesar de su creciente sofisticación técnica, están socavados por limitaciones fundamentales que abarcan todo el proceso, del dato a la estrategia. La calidad de los datos se ve comprometida por silos, el desbalance crónico de clases y un enfoque reactivo en la creación de características, vulnerable al concept drift. A su vez, los motores de detección sean sistemas de reglas o algoritmos avanzados de Machine Learning, exhiben fallos conceptuales críticos: la rigidez de los primeros y la opacidad y, sobre todo, la "visión de túnel transaccional" de los segundos.

La suma de estas debilidades culmina en un paradigma de detección sistémicamente reactivo y fragmentado. Reactivo, porque actúa cuando el daño financiero ya ocurrió, en lugar de prevenirlo en las fases iniciales del ciclo criminal. Fragmentado, porque al analizar cuentas de forma aislada, es ciego a las redes coordinadas que son la verdadera firma del crimen organizado.

Estas carencias no son meros retos técnicos a resolver con un algoritmo mejor o más datos. Son fallos de paradigma que exigen una reconceptualización del problema. Justifican inequívocamente la necesidad de explorar un nuevo modelo que sea, por diseño, holístico, proactivo y relacional. Un modelo que no se limite a clasificar transacciones, sino que entienda el ciclo de vida del mulero en sus fases y que use la potencia del análisis de grafos para desarticular las redes criminales desde su núcleo. La construcción y validación de tal modelo será el objetivo central de los siguientes capítulos de este trabajo.

CAPÍTULO 4. MODELO PROPUESTO: DETECCIÓN INTEGRADA BASADA EN EL CICLO DE VIDA DEL CLIENTE

El análisis crítico del capítulo precedente sobre las arquitecturas de detección de fraude actuales desveló limitaciones conceptuales y estructurales que merman su eficacia ante la amenaza organizada de las cuentas mula. Quedó en evidencia cómo la dependencia de datos dispersos, el crónico desbalance de clases, una ingeniería de características anclada en el pasado y, crucialmente, una "visión de túnel transaccional", dan como resultado un paradigma defensivo por naturaleza reactivo y fragmentado. Este enfoque, enfocado en clasificar eventos aislados, se muestra estructuralmente incapaz de anticipar el riesgo.

Frente a este panorama, este capítulo articula una alternativa fundamental: un cambio de paradigma que transita desde la clasificación de eventos discretos hacia el análisis de procesos continuos. Se propone un modelo de detección holístico, proactivo y contextual, con el ciclo de vida del cliente en el ecosistema financiero como eje vertebrador. La meta no es ya mejorar la precisión de un algoritmo, sino redefinir el problema en sí. La pregunta deja de ser "¿es esta transacción fraudulenta?" para convertirse en una más estratégica: "¿se ajusta el comportamiento íntegro de este cliente a las fases y patrones de una cuenta instrumentalizada para fines ilícitos?".

Para lograrlo, descompondremos el ciclo de vida del cliente en tres fases críticas (*Figura.8*): Onboarding, Operativa y Desconexión/Huida, analizando en cada una las señales, tanto explícitas como latentes, que distinguen a un actor legítimo de un mulero. Exploraremos cómo la integración de datos contextuales y conductuales en cada etapa permite construir un perfil de riesgo dinámico. Finalmente, se delineará la arquitectura funcional del modelo y se evaluará su viabilidad, demostrando que alinear la defensa con la narrativa del ciclo delictivo permite pasar de la contención de daños a una prevención real del fraude.

Figura 8. Estrategia de Detección por Fases del Ciclo de Vida.

4.1. El Ciclo de Vida como Marco Conceptual para una Detección Proactiva

La eficacia de cualquier modelo predictivo se ancla en su capacidad para abstraer la realidad de forma computacionalmente tratable y fiel al fenómeno que pretende modelar. Al enmarcar el fraude como un problema de clasificación de transacciones, los sistemas tradicionales adoptan una abstracción que, si bien es conveniente matemáticamente, ignora la dimensión procesal del crimen financiero. Una cuenta mula no debe concebirse como un estado estático, sino como la manifestación de un proceso criminal estructurado; una secuencia de acciones coordinadas que se despliegan en el tiempo. Por ello, un modelo de detección eficaz debe reflejar esta realidad procesal en su diseño.

El concepto de ciclo de vida del cliente ofrece justo este marco. En lugar de analizar interacciones como eventos aislados, este enfoque las sitúa en el contexto de la relación continua entre cliente y entidad. Dicha perspectiva trasciende la inmediatez de la transacción para evaluar la coherencia del comportamiento, contrastando acciones presentes con el historial pasado y las expectativas futuras. La operativa de un mulero, como se ha visto, sigue un ciclo predecible que pervierte el de un cliente genuino, desde la captación y apertura hasta su explotación y abandono. Adoptar este mismo ciclo como

eje de análisis sincroniza al sistema de detección con la lógica del adversario, permitiendo una identificación de anomalías más rica y contextual.

Este cambio de perspectiva rompe la "visión de túnel transaccional". Una transferencia que aisladamente no alcanzaría un umbral de riesgo significativo adquiere un nuevo sentido como parte de una secuencia: una cuenta recién abierta (onboarding), que recibe fondos de orígenes desconocidos y los dispersa de inmediato (operativa), para luego quedar inactiva (desconexión). La señal inequívoca de fraude reside en la narrativa completa, no en la transacción individual. El problema, por tanto, se transforma de una clasificación de datos a una interpretación de comportamientos.

4.2. Fase I: Onboarding – La Génesis del Riesgo y la Primera Línea de Defensa

El onboarding, o alta de cliente, es la puerta de entrada al ecosistema financiero y, en consecuencia, la oportunidad más crucial para una detección proactiva. Mientras que tradicionalmente este proceso se ha abordado desde una óptica de cumplimiento normativo para satisfacer requisitos de “Know Your Customer” (KYC) estipulados por reguladores como la EBA, un modelo de ciclo de vida lo redefine. El onboarding deja de ser una obligación burocrática para convertirse en la primera línea de defensa estratégica. El objetivo no es solo validar una identidad, sino evaluar la intencionalidad y el riesgo inherente a la nueva relación desde su concepción.

Las redes criminales explotan la eficiencia de los procesos de alta digital, diseñados para minimizar la fricción. Son conscientes de que los sistemas automatizados, aunque rápidos, son vulnerables a documentos de identidad falsificados o datos sintéticos. El modelo propuesto aquí defiende un enriquecimiento masivo del análisis en esta fase, integrando señales contextuales y conductuales que superan con creces la simple validación de un documento.

La primera capa de análisis se enfoca en el contexto tecnológico del alta. Una evaluación profunda de la inteligencia de dispositivo puede desvelar anomalías críticas. Esto abarca la huella digital del dispositivo (sistema operativo, navegador, idioma), su reputación

(¿asociado a fraude previo?) y su configuración (¿usa emulador, VPN?). Una práctica común en redes de mulas es emplear un mismo dispositivo o rango de IPs para orquestar múltiples altas, un patrón que el análisis de enlaces identifica en tiempo real. La incoherencia entre los datos del dispositivo (ej. una zona horaria que no corresponde con la IP o la dirección declarada) constituye una alerta de primer orden.

La segunda capa, inspirada en la autenticación conductual, trasciende la información introducida para enfocarse en "cómo" el usuario interactúa con el formulario de alta. La biometría del comportamiento modela patrones neuromotores únicos para cada individuo. Durante el proceso de onboarding, se capturan y analizan métricas como la velocidad y ritmo de tecleo, el movimiento del ratón o la interacción táctil. Un usuario legítimo introduciendo sus datos exhibe un patrón de familiaridad y fluidez. En cambio, un mulero con datos de terceros o un bot mostrarán patrones anómalos: tecleo inusualmente rápido y perfecto, uso extensivo de copiar y pegar en campos personales, o navegación errática. Estos patrones de "estrés" o "falta de familiaridad", como los definen algunos estudios, son potentes indicadores de que el individuo no es quien dice ser.

En la siguiente *Tabla. 3* se pueden observar una serie de variables de ejemplo más comunes utilizadas para el entrenamiento de los modelos de detección utilizando los datos de alta del cliente.

Tabla 3. Variables raw/derivadas - Onboarding

Tipo	Categoría	Variables
Original	Identificación personal	Nombre, fecha nacimiento, ID, domicilio
Original	Datos de contacto	Teléfono, email
Original	Información socioeconómica	Ocupación, ingresos, estado laboral
Original	Metadatos técnicos	IP, dispositivo, ubicación, sistema operativo

Original	Verificaciones externas	PEP, sanciones, historial de fraude
Original	Biometría	Selfie, verificación facial, huella
Derivada	Identidad/KYC	Edad, consistencia datos, flags KYC
Derivada	Contacto y reutilización	Reutilización y reputación de email/teléfono
Derivada	Dispositivo/ubicación	Distancia domicilio-IP, uso repetido, proxy/VPN
Derivada	Comportamiento en el alta	Tiempo de registro, reintentos
Derivada	Comportamiento en el alta	Patrones de digitación/ratón
Derivada	Dispositivo/ubicación	Dispositivo/IP compartidos

El proceso de alta digital, tal como explora la literatura técnica, puede modelarse como una secuencia de eventos. La aplicación de técnicas de “Process Mining” permite definir un “golden path” que representa el flujo de un cliente genuino promedio. Las desviaciones de esta ruta, como la repetición de pasos (ej. múltiples intentos fallidos de verificación), el tiempo anómalo en ciertas pantallas o una navegación no secuencial, pueden señalar una tentativa de fraude. Al tratar el “onboarding” como un log de eventos, es posible cuantificar estas desviaciones y generar una puntuación de riesgo procesal que complemente las señales de dispositivo y conducta.

La integración de estas tres capas construye un perfil de riesgo inicial rico y multidimensional. No se trata de una decisión binaria "aceptar/rechazar", sino de una puntuación dinámica que condicionará la monitorización futura. Una cuenta abierta desde

un dispositivo reputado, con comportamiento fluido y siguiendo el flujo esperado, iniciará su ciclo con riesgo bajo. En cambio, una cuenta creada desde una red anónima, con comportamiento errático y desviaciones procesales, será etiquetada con riesgo elevado desde su origen, activando una vigilancia intensificada antes de la primera transacción.

4.3. FASE II: OPERATIVA – CONTEXTUALIZACIÓN DEL COMPORTAMIENTO TRANSACCIONAL

En la fase operativa, núcleo del ciclo de vida, se manifiesta el comportamiento financiero del cliente. Es aquí donde los sistemas tradicionales concentran sus esfuerzos, analizando transacciones en busca de anomalías. El modelo propuesto, sin embargo, trasciende este enfoque al contextualizar cada transacción dentro de la narrativa continua del cliente, integrando el perfil de riesgo heredado del onboarding y enriqueciéndolo con un análisis conductual.

4.3.1. PERFILADO DINÁMICO Y DETECCIÓN DE DESVIACIONES

El análisis en esta fase se enfoca en la construcción de un perfil de comportamiento dinámico e individualizado para cada cliente. Este perfil no se fundamenta en reglas estáticas, sino en el aprendizaje continuo del patrón de normalidad de cada usuario. Se modelan múltiples dimensiones de su actividad financiera, tales como:

- Patrones monetarios: valor medio y desviación de las transacciones, distribución de importes, frecuencia y volumen de operaciones.

- Patrones temporales: horarios y días de la semana en los que el cliente suele operar.

- Patrones geográficos: ubicaciones (IPs, geolocalización) desde las que accede habitualmente.

- Patrones de contrapartida: beneficiarios y ordenantes recurrentes, países de destino/origen de fondos.

- Patrones de canal: uso preferente de la app móvil, web o cajeros.

En el siguiente apartado se muestra una tabla donde se pueden observar variables crudas y trasformadas/derivadas de ejemplo que se pueden captar y recolectar con la operativa de cada cliente. (*Tabla.4*)

Tabla 4. Variables raw/derivadas - Operativa

Tipo	Categoría	Variables
Original	Transaccionales	Monto, tipo, fecha, canal, país destino
Original	Eventos de cuenta	Inicios de sesión, cambios en cuenta
Original	Dispositivo y ubicación	IP, dispositivo, localización geográfica
Original	Estados y balances	Saldos, límites, tipo de cuenta
Original	Alertas internas	Alertas AML, flags internos
Original	Datos relacionales	Relaciones entre cuentas
Derivada	Transacciones: frecuencia y volumen	Frecuencia, ratio entradas/salidas, velocidad
Derivada	Contrapartes y destino	Nuevos destinatarios, concentración pagos
Derivada	Comportamiento de acceso	Frecuencia logins, restablecimiento clave
Derivada	Indicadores técnicos	IP nueva, dispositivo nuevo, ubicación atípica

Derivada	Análisis de ciclo transaccional	Ciclos de dinero (de A a B a C y regreso)

Una vez establecido el perfil de normalidad, el sistema no busca transacciones que superen un umbral absoluto, sino que detecta desviaciones significativas respecto al comportamiento histórico del propio cliente. Este perfilado dinámico permite una detección más precisa con menos falsos positivos. Una transferencia de 5.000€ puede ser normal para un cliente con alto poder adquisitivo, pero extremadamente anómala para un estudiante cuya cuenta nunca ha superado los 500€. Esta relativización del riesgo al comportamiento individual identifica a las cuentas mula, que exhiben patrones radicalmente distintos a los de un cliente genuino: la recepción súbita de fondos que no se corresponden con su perfil, seguida de una rápida dispersión a contrapartidas desconocidas, a menudo internacionales, dejando el saldo próximo a cero en el fenómeno conocido como "nivelación".

4.3.2. INGENIERÍA DE CARACTERÍSTICAS AVANZADA PARA LA MODELIZACIÓN CONDUCTUAL

Para materializar el perfilado dinámico, la ingeniería de características (feature engineering) se vuelve un proceso sofisticado y central, enfocado en cuantificar las desviaciones del comportamiento individual. En lugar de utilizar únicamente atributos brutos, se construyen nuevas variables que encapsulan el contexto histórico del cliente, a menudo mediante ventanas temporales deslizantes (ej. 1h, 24h, 7d, 30d).

Ejemplos de características avanzadas incluyen:

- Características de desviación monetaria: Se calcula la puntuación Z o la desviación estándar del importe de la transacción actual con respecto a la media y desviación del cliente en diferentes ventanas temporales.

- Características de frecuencia y velocidad: Se mide la frecuencia de transacciones en ventanas cortas. Variables como "número de transacciones en la última hora" o "tiempo

desde la última transacción" son críticas para detectar ráfagas de actividad anómalas, un sello distintivo de las mulas.

- Características de novedad categórica: Se crean variables binarias que indican si un atributo de la transacción es nuevo para el cliente: “NuevoDispositivo”, “NuevaIP”, “NuevoBeneficiario”, “NuevoPaisDestino”. Un cúmulo de "novedades" en una transacción eleva significativamente el riesgo.

- Características de comportamiento agregado: Se computan agregados sobre el historial reciente, como "suma total transferida en las últimas 24 horas" o "número de países distintos a los que se ha enviado dinero en la última semana". Capturan patrones a una escala mayor que la transacción individual.

Este enfoque de ingeniería de características convierte cada transacción en un rico vector de información contextualizada, permitiendo a los modelos de Machine Learning aprender no solo patrones de fraude genéricos, sino la ruptura de la normalidad individual.

4.4. FASE III: DESCONEXIÓN Y RETROALIMENTACIÓN – APRENDIZAJE CONTINUO

Frecuentemente ignorada por los sistemas de detección, la fase final del ciclo de vida, desconexión o huida representa una fuente de inteligencia de valor incalculable. Esta fase comienza cuando cesa la actividad de la cuenta mula, bien porque ha sido bloqueada por el banco, o porque la red criminal la ha "descartado" al considerarla "quemada".

El modelo propuesto resignifica esta fase, convirtiéndola de un punto final a un mecanismo de retroalimentación fundamental. Cuando un analista confirma una cuenta como mula, la etiqueta se propaga a todo su ciclo de vida registrado. El conjunto de datos asociado a esa cuenta, desde las señales del onboarding hasta el último patrón transaccional, se transforma en un ejemplo de fraude confirmado de alta calidad.

Este conjunto de datos etiquetado sirve para reentrenar y refinar continuamente los modelos de ML, permitiendo su adaptación a nuevas tácticas adversarias. Tal proceso de aprendizaje continuo es la defensa más eficaz contra el “concept drift”, el fenómeno por

el cual los modelos pierden eficacia a medida que el fraude evoluciona. Si los criminales desarrollan una nueva técnica, en cuanto la primera red que la utiliza es identificada, el sistema aprende el nuevo patrón y ajusta sus parámetros para detectarlo proactivamente en el futuro.

Adicionalmente, el análisis de la fase de desconexión en sí misma puede revelar patrones. Las cuentas mula, una vez explotadas, a menudo quedan en estado durmiente con saldo cero o son abandonadas abruptamente. El seguimiento de estos tipos de cuentas y su vinculación con otras puede ayudar a identificar la escala de una operación neutralizada y descubrir otros nodos aún activos.

4.5. Gobernanza, Interpretabilidad y Mantenimiento del Modelo

Un modelo de esta sofisticación exige, más allá de su desarrollo técnico, un marco de gobernanza robusto que asegure su interpretabilidad, justicia y mantenimiento. La confianza en el sistema, tanto de analistas internos como de reguladores, depende de la capacidad de entender, validar y auditar sus decisiones.

4.5.1. El Desafío de la "Caja Negra" y la Necesidad de la IA Explicable (XAI)

La opacidad de los modelos de Machine Learning más potentes, como XGBoost o las redes neuronales, su naturaleza de "caja negra", es un obstáculo crítico en el sector financiero por razones de cumplimiento (ej. el derecho a una explicación bajo GDPR), validación interna y eficacia operativa.

Para solventar esto, el modelo debe integrar técnicas de IA Explicable (XAI). Herramientas como LIME o, preferiblemente," SHAP", deben aplicarse a cada predicción de riesgo. Estas técnicas desglosan la puntuación, indicando qué características específicas contribuyeron más a la decisión y en qué medida.

Esta capacidad de explicación transforma la alerta de una puntuación a una narrativa comprensible para el analista ("El riesgo es alto porque el cliente opera de noche, desde una IP nunca vista, y el importe es 10 veces su media histórica"). Esto no solo agiliza una investigación más dirigida, sino que también proporciona una base sólida para justificar las acciones tomadas ante clientes y reguladores.

4.5.2. Marco de Gobernanza del Modelo: Monitorización y Lucha contra el Drift

Los modelos de detección de fraude no son soluciones estáticas; son sistemas dinámicos que deben adaptarse a un entorno cambiante. El “concept drift” (evolución de las tácticas de fraude) y el “data drift” (cambios en la distribución de los datos) pueden degradar rápidamente el rendimiento del modelo si no se gestionan activamente.

Se debe establecer un marco de gobernanza que incluya:

- Monitorización Continua del Rendimiento: El rendimiento del modelo (precisión, recall) debe ser monitorizado de manera continua. Es necesario establecer umbrales de alerta que se activen si el rendimiento cae por debajo de un nivel aceptable.

- Detección de Drift: Implementar algoritmos estadísticos para detectar cambios significativos en la distribución de las características de entrada y en la relación entre estas y las predicciones, como señal temprana de obsolescencia.

- Bucle de Retroalimentación Humana: Crear un flujo de trabajo para que las conclusiones de los analistas (confirmación de alertas) se registren y se usen para re-etiquetar los datos.

- Estrategia de Reentrenamiento Definida: Establecer una política clara sobre cuándo y cómo reentrenar el modelo, ya sea de forma periódica o activada por la detección de drift o una caída del rendimiento, con validación rigurosa antes de su despliegue.

Este marco de gobernanza asegura que el modelo no solo sea eficaz en su lanzamiento, sino que mantenga su relevancia y precisión a lo largo del tiempo, convirtiéndolo en una defensa sostenible.

4.6. ARQUITECTURA FUNCIONAL Y VIABILIDAD DEL MODELO PROPUESTO

Materializar un modelo de detección basado en el ciclo de vida descansa sobre una arquitectura de datos y sistemas robusta, capaz de integrar información de múltiples fuentes en tiempo real, ejecutar complejos procesos de ingeniería de características y aplicar modelos de clasificación.

Técnicamente, su construcción representa un desafío significativo, aunque factible con las tecnologías actuales de big data y ML. El reto principal reside en la integración de datos de silos departamentales y en la garantía de una baja latencia en todo el flujo de trabajo. Operativamente, su implementación exigiría una estrecha colaboración entre los equipos de datos, fraude, compliance y tecnología.

Una de las ventajas clave del modelo es su capacidad para reducir la carga de trabajo de los analistas. Al proporcionar alertas de mayor calidad y contextualizadas con una narrativa clara, permite a los equipos de investigación enfocarse en casos de mayor impacto y en el análisis de redes, en lugar de revisar un volumen masivo de falsos positivos.

El modelo de detección basado en el ciclo de vida del cliente supone una evolución fundamental en la lucha contra el fraude con cuentas mula. Al abandonar la visión de túnel transaccional por un enfoque holístico que analiza el comportamiento del cliente a lo largo del onboarding, la operativa y la desconexión, el sistema adquiere una capacidad sin precedentes para contextualizar el riesgo.

La integración de señales de inteligencia de dispositivo, biometría conductual y una sofisticada ingeniería de características permite al modelo no solo reaccionar, sino anticipar el riesgo desde la génesis de la relación. La incorporación de un marco de gobernanza robusto, con énfasis en la IA Explicable y la lucha contra el concept drift, asegura que el modelo sea transparente, justo y sostenible.

Este enfoque proactivo y contextual, sustentado por un bucle de retroalimentación continuo, lo posiciona como una defensa resiliente frente a un adversario en constante evolución. Este capítulo ha sentado sus bases conceptuales y arquitectónicas. Su construcción no es meramente un ejercicio técnico, sino un replanteamiento estratégico que prepara el terreno para explorar, en capítulos posteriores, su extensión hacia el análisis de redes criminales complejas.

CAPÍTULO 5. ANÁLISIS DE GRAFOS Y APRENDIZAJE AUTOMÁTICO COMO CATALIZADORES DEL MODELO

Los capítulos precedentes de este trabajo han dejado claro las limitaciones inherentes a los paradigmas de detección de fraude financiero contemporáneos. Ha quedado de manifiesto que, pese a una creciente sofisticación algorítmica, los enfoques centrados en el análisis transaccional aislado sufren de una "visión de túnel". Dicha perspectiva, reactiva y fragmentada por naturaleza, es estructuralmente incapaz de anticipar y desarticular las operaciones coordinadas de las redes de cuentas mula, que hoy son la espina dorsal del blanqueo de capitales digital. La dependencia de datos en silos, junto al crónico desafío del desbalance de clases y una ingeniería de características anclada a patrones de fraude ya conocidos, condenan a los sistemas actuales a una perpetua carrera por detrás del adversario.

Frente a ello, este capítulo propone un giro paradigmático fundamental: una metodología holística, proactiva y relacional que trasciende la simple clasificación de eventos discretos para abrazar el análisis de procesos criminales en su totalidad. El objetivo es catalizar una nueva generación de sistemas de detección mediante la sinergia de dos pilares tecnológicos: el análisis de grafos. Se defenderá la idea de que representar las interacciones financieras como un grafo no es una mera elección técnica, sino un imperativo conceptual para poder desvelar las estructuras ocultas del fraude organizado.

El eje de esta propuesta lo constituye un modelo basado en grafos. Este enfoque dinámico modela la evolución del comportamiento de un cliente no como una serie de eventos inconexos, sino como una trayectoria continua a través de diferentes fases. Se detallará cómo esta arquitectura permite una detección más temprana y precisa, una reducción significativa de falsos positivos y la capacidad de identificar y desmantelar redes completas. Por último, se abordarán los aspectos más críticos para su implementación en

un entorno productivo, tales como la explicabilidad del modelo (XAI), su gobernanza algorítmica y las estrategias para asegurar su sostenibilidad frente al fenómeno del concept drift.

5.1. FUNDAMENTOS DEL ANÁLISIS DE GRAFOS EN LA DETECCIÓN DE FRAUDE

La transición del enfoque tabular al basado en grafos no es una mera actualización técnica; representa, de hecho, la evolución conceptual más significativa en la lucha moderna contra el fraude financiero. Mientras que los modelos tradicionales, operando sobre datos en filas y columnas, son eficientes para analizar atributos individuales, fallan sistemáticamente al intentar capturar la riqueza de las relaciones. El fraude, en particular el orquestado mediante redes de mulas, no es un fenómeno de atributos, sino de conexiones. El análisis de grafos se erige así no como una simple herramienta, sino como el lenguaje natural para describir y comprender estas actividades ilícitas.

Matemáticamente, un grafo es una estructura de nodos (vértices) y aristas (ejes) que los conectan. En el dominio financiero, los nodos pueden representar una gran diversidad de entidades: clientes, cuentas, dispositivos, direcciones IP, teléfonos, comerciantes o transacciones. Las aristas, a su vez, modelan las relaciones explícitas e implícitas entre ellos: una transacción, la titularidad. el uso de un dispositivo o un dato compartido. Este modelo de datos, nativamente conectado, supera la rigidez de las bases de datos relacionales y elimina la necesidad de costosas operaciones de JOIN para reconstruir relaciones que en el grafo ya existen.

La anatomía de un grafo queda desglosada en la *Figura. 9*, que clarifica sus componentes. En ella, los nodos operan como las entidades fundamentales y las aristas como los vínculos que materializan las relaciones entre ellos. Mención aparte merecen las aristas ponderadas: conexiones que portan información cuantitativa, como el importe de una transacción.

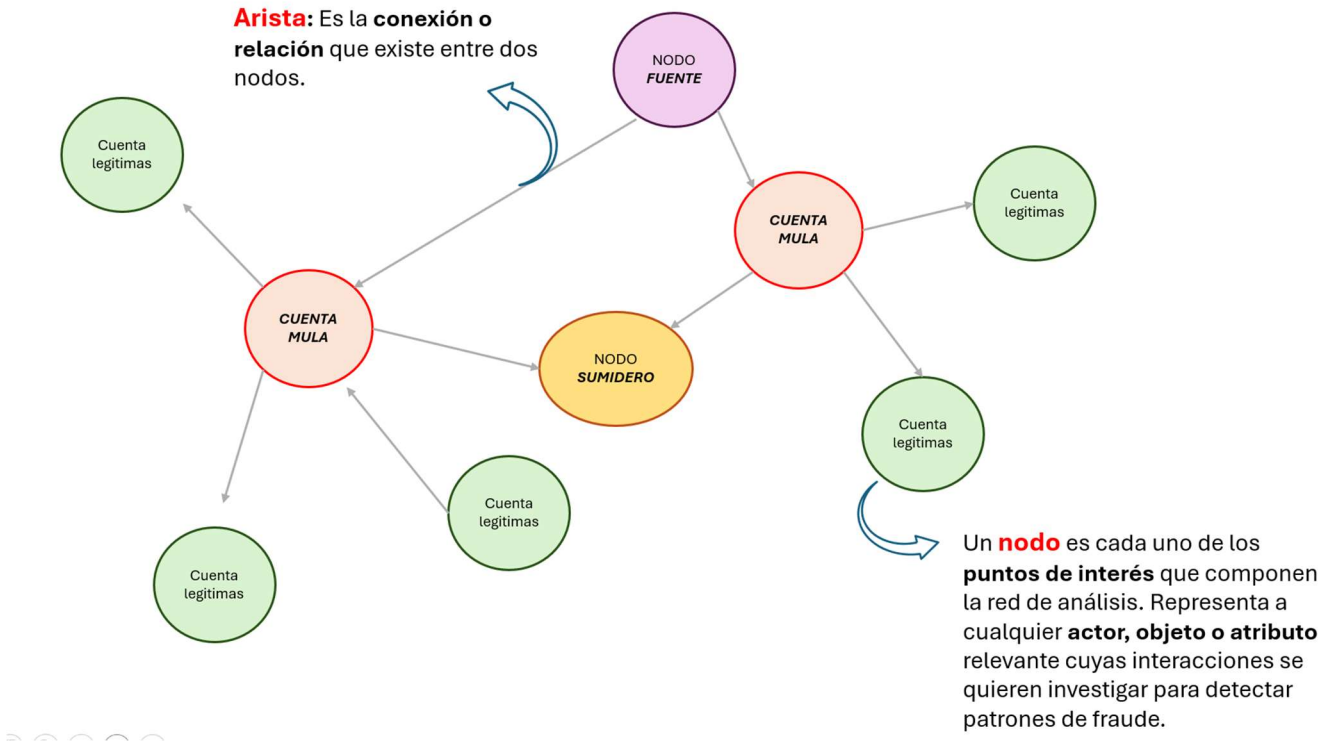

Figura 9. Componentes Fundamentales de un Grafo: Nodos, Aristas.

El valor del análisis de grafos reside en su potencial para desvelar patrones invisibles desde una perspectiva aislada. Su capacidad para el descubrimiento de relaciones ocultas es clave para identificar intermediarios y conexiones indirectas, cruciales en el blanqueo. Dos cuentas sin relación aparente pueden vincularse a través de un mismo dispositivo, una IP o una cadena de transacciones intermediarias (mulas). El análisis de grafos facilita, además, la detección de comunidades, grupos de nodos densamente conectados.

Algoritmos como Louvain Modularity o Weakly Connected Components (WCC) permiten identificar estos clústeres, que a menudo se corresponden con anillos de fraude coordinados.

Esta metodología permite también un profundo análisis de centralidad para cuantificar la importancia de cada actor. Métricas como el "Grado de Centralidad" miden conexiones directas e identifican cuentas que agregan o dispersan fondos; la Centralidad de

Intermediación detecta nodos "puente", rol típico de las mulas; y algoritmos como “PageRank” miden la influencia de un nodo según la de sus vecinos, ayudando a localizar coordinadores de redes. Los grafos son, además, excepcionales para identificar patrones topológicos de colusión, como estructuras en estrella o ciclos de transacciones, característicos de tácticas de blanqueo difíciles de detectar con otros métodos.

La construcción de un grafo de conocimiento (Knowledge Graph - KG) enriquece todavía más este análisis al integrar datos heterogéneos con semántica explícita. Un KG no solo conecta clientes y transacciones; puede incorporar datos no estructurados o contextuales como la reputación de un dispositivo. Esta visión 360°, centralizada en un único modelo, provee el contexto necesario para diferenciar anomalías genuinas de coincidencias benignas, reduciendo así los falsos positivos.

5.2. DETECCIÓN DE REDES MULEADAS MEDIANTE TOPOLOGÍAS DE GRAFO

El análisis de grafos se vuelve especialmente poderoso al aplicarse al problema concreto de las redes de mulas. Estas redes no son una colección aleatoria de cuentas fraudulentas; son estructuras organizadas con roles y topologías definidas, diseñadas para ofuscar al máximo el rastro del dinero. La habilidad para modelar y analizar estas topologías en tiempo real es lo que marca el paso de una detección de mulas individuales a un desmantelamiento de operaciones de blanqueo a escala industrial.

Dentro de una red de blanqueo, los nodos adoptan roles funcionales muy específicos. Primero, se identifican los nodos fuente, las cuentas de las víctimas desde donde parte el dinero ilícito, caracterizadas por un alto volumen de transacciones salientes hacia nodos sin relación previa, como se observa en la *Figura.10*. Después se encuentran los nodos intermediarios, las propias cuentas mula, que forman el cuerpo de la red y cuya función es recibir y transferir fondos velozmente para añadir capas de complejidad, distinguiéndose por un patrón de "entrada-salida" casi inmediato.

Finalmente, el dinero alcanza los nodos sumidero, destino final de los fondos "lavados", que se caracterizan por recibir dinero de múltiples intermediarios sin una actividad de

salida proporcional y que a menudo son exchanges de criptomonedas o cuentas en jurisdicciones de baja regulación.

Estas funciones se organizan en topologías de red que se repiten, cada una optimizada para un fin. Una de las más comunes es la topología en estrella, donde múltiples cuentas fuente envían fondos a una mula central que, a su vez, los dispersa a múltiples destinos. En métricas de grafo, el nodo central exhibirá un alto grado de centralidad e intermediación. Otra estructura frecuente es la topología en cadena: los fondos se mueven secuencialmente a través de varias mulas para alargar y complicar el rastreo, táctica identificable con algoritmos de búsqueda de caminos. También pueden surgir topologías cíclicas, si bien son menos comunes, para circular dinero en un grupo cerrado de cómplices.

La *Figura.10* esquematiza las dos topologías más prevalentes en las redes de blanqueo. La primera, una estructura vertical en cadena, detalla cómo un emisor único fracciona los fondos entre múltiples intermediarios para su posterior consolidación en un receptor. La segunda, una secuencia horizontal, ilustra el tránsito de fondos a través de una cadena de intermediarios hasta su destino final. Ambas tácticas comparten el objetivo de ofuscar el origen del dinero mediante la estratificación.

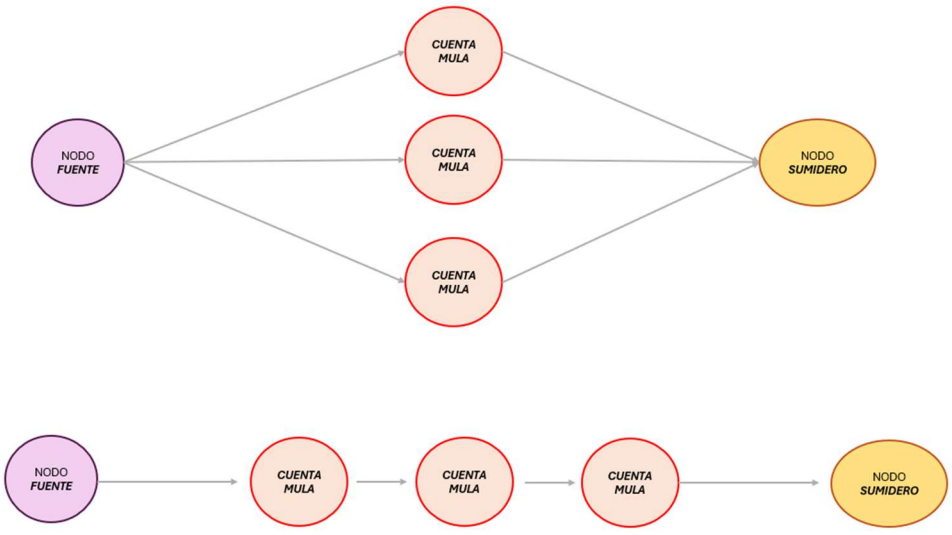

Figura 10. Topologías Comunes en Redes de Blanqueo de Capitales .

Sobre este grafo, la *Figura. 11* visualiza patrones de alerta concretos: el de recolección y dispersión (gather-scatter) se manifiesta como una nítida estructura en estrella, mientras

que los patrones cíclicos, usados para crear confusión, son también identificables. La detección de estas geometrías en un mar de transacciones resulta decisiva para identificar eficaz y tempranamente el blanqueo.

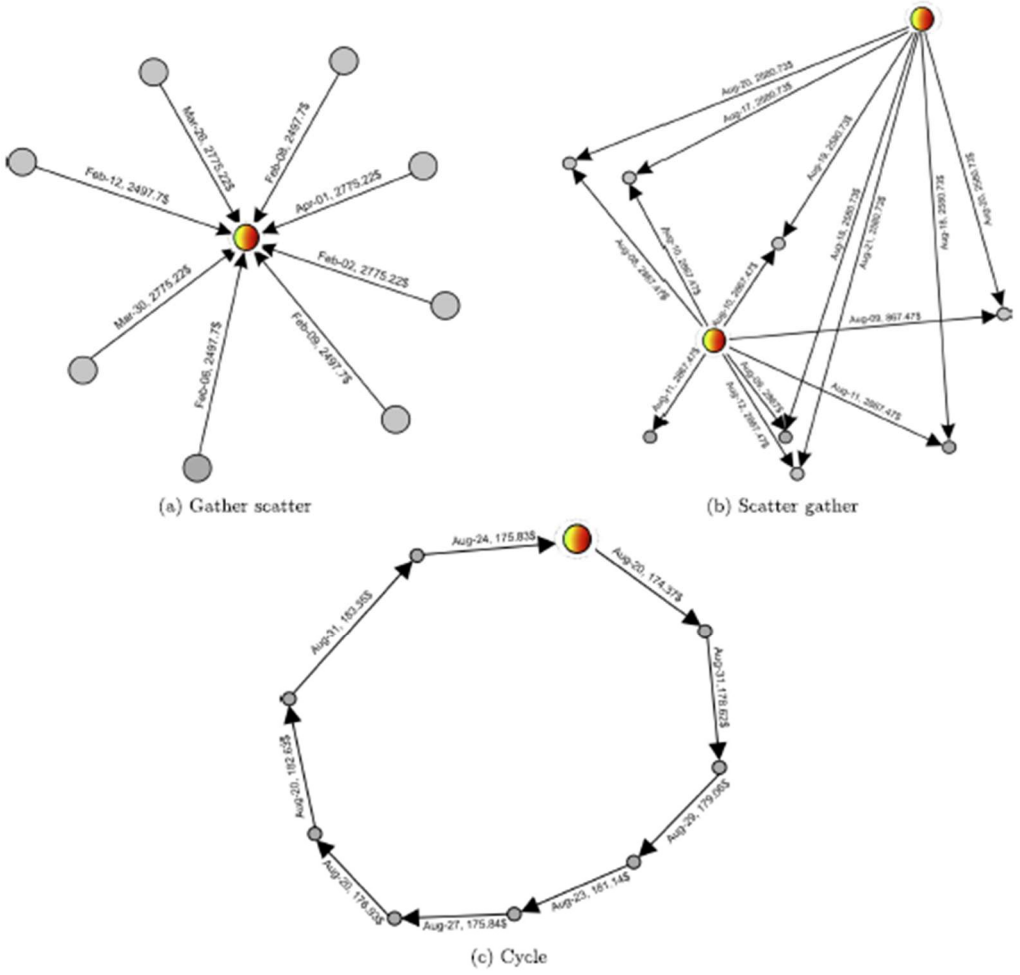

Figura 11. Visualización de Patrones de Blanqueo en el Grafo de AML [1]

Para que estos análisis topológicos resulten efectivos, es crucial enriquecer el grafo con características temporales y contextuales. La combinación de métricas estructurales con atributos de nodos y aristas, importe, frecuencia, antigüedad de la cuenta, reputación del dispositivo, dota a los modelos de aprendizaje automático de un conjunto de características con una riqueza predictiva inalcanzable para los sistemas tradicionales.

5.3. EXPLICABILIDAD (XAI), GOBERNANZA Y SOSTENIBILIDAD DEL MODELO

El despliegue de modelos avanzados en el entorno financiero no es solo un desafío técnico; es, ante todo, un reto de gobernanza, confianza y transparencia. Un modelo puede ser muy preciso, pero si sus decisiones son inescrutables su utilidad práctica queda severamente limitada. Para superar esta opacidad, es fundamental integrar técnicas de IA Explicable (XAI) en el flujo de trabajo. Los valores SHAP adaptados a grafos permiten diseccionar una predicción, señalando qué vecinos o características influyeron más en ella. Este nivel de detalle transforma una alerta opaca en una narrativa de riesgo procesable, acelerando la investigación y aportando la justificación exigida por la normativa.

La implementación de este tipo de modelo exige, además, un marco de gobernanza robusto que defina la propiedad, validación y responsabilidad de sus decisiones. Esto implica la auditoría regular de sesgos, la monitorización continua de su rendimiento y protocolos claros para que los analistas humanos puedan revisar y anular decisiones del modelo, creando así un bucle de retroalimentación esencial.

Finalmente, la sostenibilidad del sistema descansa en su capacidad para adaptarse al concept drift, la constante evolución de las tácticas de fraude. La arquitectura basada en grafos, sumada a un bucle de aprendizaje continuo, es la defensa más sólida contra este fenómeno. Al detectar cambios en la distribución de datos, recibir retroalimentación de analistas sobre casos confirmados y reentrenar adaptativamente el modelo, el sistema asimila nuevos patrones de ataque. Este marco de gobernanza y mantenimiento convierte el modelo en un sistema defensivo vivo y resiliente, que aprende del adversario y asegura su eficacia a largo plazo.

Este capítulo ha presentado una metodología avanzada para la detección de cuentas mula, fundamentada en la sinergia de análisis de grafos y aprendizaje automático. Se ha demostrado que el salto conceptual desde el análisis transaccional hacia una visión relacional es indispensable para combatir las redes de fraude organizadas. La representación de interacciones financieras como un grafo permite desvelar topologías de colusión, identificar actores clave con métricas de centralidad y descubrir relaciones ocultas, invisibles para los modelos convencionales.

Se ha subrayado, además, que la viabilidad de un sistema así depende de un sólido marco de gobernanza que garantice la explicabilidad con técnicas de XAI y asegure su sostenibilidad combatiendo el concept drift mediante aprendizaje continuo. La metodología propuesta no busca una simple mejora incremental de métricas; redefine el paradigma de la lucha contra el fraude. Al pasar de clasificar eventos a interpretar procesos y de analizar nodos aislados a desarticular redes completas, se ofrece a las instituciones una herramienta estratégica para transitar desde la contención de daños hacia una prevención real y anticipación de amenazas.

CAPÍTULO 6. CONCLUSIONES

Este Trabajo Fin de Máster ha abordado el fenómeno de las cuentas mula no como una externalidad de la digitalización, sino como una amenaza estructural en la intersección del fraude, el blanqueo y la ingeniería social, cuya proliferación socava los cimientos del ecosistema bancario. La investigación ha transitado desde un diagnóstico crítico de las arquitecturas de detección vigentes hasta la formulación de un paradigma alternativo de carácter holístico y proactivo. Este capítulo final no es una mera recapitulación, sino una síntesis reflexiva sobre las implicaciones estratégicas de los hallazgos, sus desafíos y las futuras líneas de investigación que se abren, proyectando así el conocimiento generado hacia la praxis del sector.

El análisis acometido ha revelado una paradoja fundamental: mientras la industria invierte en algoritmos de aprendizaje automático cada vez más sofisticados, su paradigma operativo permanece anclado a una "visión de túnel transaccional". Los sistemas actuales, aun siendo potentes, están diseñados para clasificar eventos discretos una transferencia, un inicio de sesión, juzgando cada acción de forma aislada. Este enfoque, por diseño, es reactivo. La alarma salta cuando el dinero ya está en movimiento y la cadena de blanqueo se ha iniciado, convirtiendo la lucha en una costosa labor de contención de daños en lugar de una prevención genuina.

La propuesta central de este trabajo ha sido desmantelar dicho paradigma para sustituirlo por un modelo integrado, cuyo eje es el ciclo de vida del cliente. Este cambio de perspectiva es radical. La pregunta deja de ser "¿es esta transacción una mula?" para transformarse en "¿se corresponde la trayectoria de comportamiento de esta cuenta, desde su origen, con la narrativa de una instrumentalización ilícita?". Al descomponer la relación del cliente en fases Onboarding, Operativa y Desconexión, el modelo contextualiza cada interacción. Se ha demostrado teóricamente cómo integrar señales de riesgo desde la génesis de la cuenta (inteligencia de dispositivo, biometría conductual) permite construir un perfil de riesgo proactivo. Este enriquecimiento se proyecta sobre la fase operativa, donde el análisis detecta desviaciones respecto a la normalidad individualizada de cada cliente, identificando las sutiles señales que delatan a una cuenta

mula: la recepción de fondos inconsistentes seguida de una dispersión casi inmediata. Finalmente, la fase de desconexión se resignifica como un bucle de retroalimentación para el reentrenamiento continuo de los modelos, dotando al sistema de una defensa adaptativa frente al concept drift.

Esta transición se justifica por las debilidades estructurales de los sistemas actuales. Los silos de información fragmentan la visión del cliente. El enfoque propuesto no se limita a balancear datos; enriquece el contexto para que la cuenta mula sea visible desde su creación. La dependencia de una ingeniería de características anclada en fraudes pasados condena a los sistemas a una perpetua carrera por detrás de un adversario adaptativo. El modelo de ciclo de vida, en cambio, se centra en detectar la ruptura de la normalidad conductual, una señal que puede preceder a nuevas tipologías de fraude.

Sin embargo, la vulnerabilidad más crítica que las redes criminales explotan es la ceguera ante las estructuras coordinadas. El crimen organizado no opera con muleros aislados. Aquí, la introducción del análisis de grafos no es una herramienta adicional, sino el catalizador del salto cualitativo. Representar las interacciones como un grafo de nodos (clientes, dispositivos) y aristas (transacciones) es el lenguaje natural para describir el fraude organizado. Este enfoque desvela las topologías de colusión, identifica a los actores clave y descubre las relaciones ocultas. La propuesta de un Grafo de Ciclo de Vida Temporal, trasciende el análisis estático para modelar la evolución dinámica de estas redes, permitiendo pasar de la identificación de nodos al desmantelamiento de la infraestructura delictiva.

La viabilidad de un modelo de esta envergadura se enfrenta a desafíos prácticos. Técnicamente, exige una arquitectura de datos moderna capaz de romper los silos y procesar grafos a gran escala. Regulatoriamente, el cruce masivo de datos choca con normativas de privacidad como el GDPR. Tecnologías emergentes de preservación de la privacidad (Privacy-Enhancing Technologies), como el aprendizaje federado, ofrecen vías prometedoras para una colaboración interbancaria respetuosa con la ley. Éticamente, un modelo que analiza el comportamiento de forma tan granular exige una gobernanza impecable para prevenir sesgos. La opacidad de los modelos más potentes debe mitigarse obligatoriamente mediante técnicas de IA Explicable (XAI), como SHAP, que traducen las decisiones algorítmicas en narrativas comprensibles, garantizando la transparencia y la confianza.

Este trabajo, además, se alinea con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas. Al proponer herramientas contra los flujos financieros ilícitos, contribuye directamente al *ODS 16* (Paz, justicia e instituciones sólidas). Un sistema financiero más seguro es, a su vez, pilar para el *ODS 8* (Trabajo decente y crecimiento económico), al fomentar un entorno de confianza para la inversión. Finalmente, el propio diseño del modelo impulsa la innovación y la creación de infraestructuras tecnológicas resilientes, en plena consonancia con el *ODS 9* (Industria, innovación e infraestructura). La lucha contra las cuentas mula trasciende así la gestión de riesgos para convertirse en un factor de sostenibilidad y desarrollo.

En última instancia, este Trabajo Fin de Máster ha buscado articular una respuesta a una de las amenazas más corrosivas de la era digital. La digitalización ha industrializado el crimen, y las cuentas mula son el engranaje clave de esta nueva maquinaria. Combatirlas eficazmente no es solo un imperativo de negocio ni una mera obligación de cumplimiento. Es una responsabilidad fundamental para proteger la integridad del sistema financiero y el tejido social que depende de él. El modelo propuesto es una hoja de ruta para transitar desde una postura defensiva hacia una estrategia de anticipación y desarticulación. Es un argumento a favor de un cambio de paradigma que ponga la inteligencia contextual y el análisis relacional en el epicentro de la defensa. Se trata, en definitiva, de una contribución conceptual para restaurar y fortalecer la confianza, ese activo intangible pero insustituible, en un sistema financiero que debe ser, ante todo, un motor de progreso seguro y equitativo.

Los hallazgos y propuestas de este estudio trazan, pues, una hoja de ruta estratégica. Su fin es guiar a las entidades financieras en su transición desde un enfoque reactivo y fragmentado hacia uno proactivo, relacional y robusto. Pasar de la idea a la implementación exigirá voluntad institucional e inversión tecnológica. Pero, sobre todo, demandará una colaboración fluida entre negocio y tecnología, unidos bajo una misión compartida: blindar la confianza del cliente en la era de la banca digital.

CAPÍTULO 7. TRABAJOS FUTUROS

Esta sección final parte del alcance original del trabajo para proponer líneas concretas de evolución. El objetivo es conectar la visión estratégica con un futuro desarrollo técnico y multidisciplinar.

Este trabajo ha presentado una propuesta conceptual para transformar el paradigma de detección de cuentas mula. Articula una solución holística y proactiva, anclada en el ciclo de vida del cliente y potenciada por análisis de grafos. El enfoque se ha mantenido, ciertamente, en un plano teórico y estratégico. Esta decisión obedece a la naturaleza académica del proyecto y a las limitaciones de acceso a datos confidenciales. Aun así, su diseño fundamental apunta directamente hacia una aplicación práctica, traducible en una solución técnica plenamente implementada.

La línea futura prioritaria es, por tanto, evolucionar este modelo hacia un producto de analítica avanzada que sea implementable. Esto implica desarrollar una arquitectura de referencia completa.

Para ello, será imprescindible constituir un equipo multidisciplinar. Este debe unir perfiles de ciencia e ingeniería de datos, ciberseguridad, arquitectura tecnológica, cumplimiento normativo y expertos de negocio. Un equipo así permitiría abordar la puesta en producción del modelo desde una perspectiva verdaderamente integral, donde visión estratégica y técnica confluyan desde el inicio. Este documento sirve, en consecuencia, como una propuesta fundacional. Una base sólida sobre la cual edificar esa futura y necesaria colaboración transversal.

Algunas líneas concretas de evolución incluyen:

- Validar las hipótesis del modelo mediante prototipos técnicos en entornos controlados (sandbox) con datos sintéticos.
- Diseñar una arquitectura tecnológica escalable que incluya orquestadores de pipelines, plataformas de grafos (Neo4j, TigerGraph), motores de scoring y robustos mecanismos de auditoría y gobernanza bajo criterios de XAI.

- Definir KPIs de rendimiento y de negocio para medir el impacto real de la solución: tasa de falsos positivos, recuperación de fondos, reducción de pérdidas y eficiencia operacional.

CAPITULO 8. BIBLIOGRAFÍA

[1] Karim, M.R.; Hermsen, F.; Chala, S.A.; De Perthuis, P.; Mandal, A. "Scalable Semi-Supervised Graph Learning Techniques for Anti Money Laundering". IEEE Access, vol. 12, pp. 50012-50029 , Abril, 2024 . http://dx.doi.org/10.1109/ACCESS.2024.3383784.

[2] Nasdaq Verafin. "Financial Crime Insights: Europe". Marzo, 2025.

[3] Grupo de Acción Financiera Internacional (GAFI). "Las Recomendaciones del GAFI". GAFI/FATF, Febrero 2012 (actualizado a junio de 2024).

[4] Grupo de Acción Financiera Internacional (GAFI). "Guía para un enfoque basado en riesgo para Activos Virtuales y Proveedores de Servicios de Activos Virtuales". GAFI/FATF, Octubre 2021.

[5] Parlamento Europeo y Consejo de la Unión Europea. "Directiva (UE) 2018/843 del 30 de mayo de 2018 por la que se modifica la Directiva (UE) 2015/849 relativa a la prevención de la utilización del sistema financiero para el blanqueo de capitales o la financiación del terrorismo". Diario Oficial de la Unión Europea, L 156/43, 19 de junio de 2018.

[6] Jefatura del Estado. "Ley 10/2010, de 28 de abril, de prevención del blanqueo de capitales y de la financiación del terrorismo". Boletín Oficial del Estado, núm. 103, de 29 de abril de 2010.

[7] Tesoro Público. "El Sepblac crea un nuevo canal para que la banca informe de las 'cuentas mula'". Cinco Días - El País, 29 de febrero de 2024.

[8] European Banking Federation. "#DontBeaMule - European Money Mule Action". Disponible en: https://www.ebf.eu/priorities/cybersecurity/dontbe-a-mule/.

[9] Pozzolo, A. D.; Caelen, O.; Le Borgne, Y-A.; Waterschoot, S.; Bontempi, G. "Calibrating Probability with Undersampling for Unbalanced Classification". En IEEE Symposium Series on Computational Intelligence, 2015, pp. 130-137.

[10] Soria, J.; Loayza, R.; Segura, L. "Machine Learning Models for Money Laundering Detection in Financial Institutions. A Systematic Literature Review". 22nd LACCEI International Multi-Conference for Engineering, Education, and Technology, Julio, 2024.

[11] Arevalo, B. C. "Money Mules: Facilitators of financial crime. An explorative research on money mules". Universidad de Utrecht, Junio 2015.

[12] Aston, M.; McCombie, S.; Reardon, B.; Watters, P. “A Preliminary Profiling of Internet Money Mules: An Australian Perspective”. Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, 2009, pp. 15-30.

[13] Kaung Wai Thar; Thinn Thinn Wai. “Machine Learning Based Predictive Modelling for Fraud Detection in Digital Banking”. 2024 IEEE Conference on Computer Applications (ICCA), 2024.

[14] Nicholls, J.; Kuppa, A.; Le-Khac, N-A. “Financial Cybercrime: A Comprehensive Survey of Deep Learning Approaches to Tackle the Evolving Financial Crime Landscape”, vol. 9, 2021.

[15] Alonso Fernández, D.; Álvarez Sánchez, A. "Año nuevo, mismos ataques". Radar: El magazine de ciberseguridad, nº 99, Febrero 2025, pp. 4-5.

[16] Camilo da Silva, M.; Marques Tavares, G.; Gritti, M. C.; Ceravolo, P.; Barbon Junior, S. "Using Process Mining to Reduce Fraud in Digital Onboarding". FinTech, vol. 2, nº 9, 2023, pp. 120-137.

[17] Cookman, D. "European approach to remote customer onboarding solutions". Journal of Money Laundering Control, vol. 26, nº 7, 2023, pp. 213-223.

[18] García Lorente, F. J. "El fraude bancario en la era digital: retos y estrategias de la ciberseguridad". Radar: El magazine de ciberseguridad, nº 99, Febrero 2025, pp. 2-3.

[19] Manwani, P. "AI-Enhanced Customer Authentication and Onboarding". International Journal of Trend in Scientific Research and Development (IJTSRD), vol. 9, nº 2, Abril 2025.

[20] Marín, G. "Altos estándares de ciberseguridad". Radar: El magazine de ciberseguridad, nº 99, Febrero 2025, pp. 9-10.

[21] Morbiato, F. "Algoritmi di Machine Learning a confronto per la rilevazione delle frodi finanziarie". Università degli studi di Padova, 19 de noviembre de 2024.

[22] Ravaglia, A. "Riconoscimento di frodi attraverso la modellazione del comportamento degli utenti". Università di Bologna, 2020.

[23] Romero Gutiérrez, A. "Medidas preventivas contra las estafas". Radar: El magazine de ciberseguridad, nº 99, Febrero 2025, pp. 6-8.

[24] Johora, F. T.; Hasan, R.; Akter, J.; Farabi, S. F.; Mahmud, M. A. A. "AI-POWERED FRAUD DETECTION IN BANKING: SAFEGUARDING FINANCIAL TRANSACTIONS". The American Journal of Management and Economics Innovations, vol. 6, nº 6, 2024, pp. 8-22.

[25] Labanca, D.; Primerano, L.; Markland-Montgomery, M.; Polino, M.; Carminati, M.; Zanero, S. "Amaretto: An Active Learning Framework for Money Laundering Detection". IEEE Access, vol. 10, 2022, pp. 41720-41739.

[26] Prabha, M.; Sharmin, S.; Khatoon, R.; Imran, M. A. U.; Mohammad, N. "COMBATING BANKING FRAUD WITH IT: INTEGRATING MACHINE LEARNING AND DATA ANALYTICS". The American Journal of Management and Economics Innovations, vol. 6, nº 7, 2024, pp. 39-56.

[27] Abdul Rani, M. I.; Zolkaflil, S.; Syed Mustapha Nazri, S. N. F. "THE MONEY MULE RED FLAGS IN ANTI-MONEY LAUNDERING TRANSACTION MONITORING INVESTIGATION". International Journal of Business and Economy, vol. 4, nº 1, 2022, pp. 150-163.

[28] Domashova, J.; Mikhailina, N. "Usage of machine learning methods for early detection of money laundering schemes". Procedia Computer Science, vol. 190, 2021, pp. 184-192.

[29] Banu, S. R.; Chowdhary, H.; Gongada, T. N.; Sabareesh R; Santosh, K.; Muthuperumal, S. "Financial Fraud Detection Using Hybrid Convolutional and Recurrent Neural Networks: An Analysis of Unstructured Data in Banking". 2024 10th International Conference on Communication and Signal Processing (ICCSP), 2024.

[30] da Silva, M. C.; Tavares, G. M.; Gritti, M. C.; Ceravolo, P.; Barbon Junior, S. "Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems". Procedia Computer Science, vol. 198, 2022.

[31] Anandhabalaji, V.; Babu, M.; Brintha, R. "Energy consumption by cryptocurrency: A bibliometric analysis revealing research trends and insights". Energy Nexus, vol. 13, 2024.

[32] "Accelerate Fraud Detection With a Graph Database & Machine Learning Platform". TigerGraph, 2022.

[33] D. S. Santolaya, J. Chaquet, S. Basaldúa, M. G. Romero, and F. M. Moreno, "Exploring graph analytics with mercury-graph," BBVA AI Factory Blog, Jan. 20, 2025. Disponible en: https://www.bbvaaifactory.com/graph-analytics-with-mercury-graph/

[34] Han, J.; Hu, H.; Ai, Y.; Wang, X.; Tian, Y.; Zhao, Z. "FFDM-GNNA: Financial Fraud Detection Model using Graph Neural Network with Attention Mechanism". 2023 2nd International Conference on Image Processing and Media Computing (ICIPMC), 2023.

[35] Addo, M-J.; Salmu, S.; Otoo, D. J-A. "Graph Neural Networks for Financial Fraud Detection: A Review". Journal of Electrical and Computer Engineering, vol. 2023, 2023.

[36] Bao, W.; Liu, J.; Zhou, X.; Liu, S.; Wang, T.; Hu, G. "TGFFD: A Two-Stream Graph Neural Network for Financial Fraud Detection Based on Graph Convolution and Wavelet Analysis". IEEE Transactions on Neural Networks and Learning Systems, 2024.

[37] Castaldi, G. "L'adeguata verifica della clientela bancaria e il recente provvedimento della Banca d'Italia". Bancaria, n. 10/2013, Octubre, 2013.

[38] Aprigliano, V.; Ardizzi, G.; Cassetta, A.; Cavallero, A.; Emiliozzi, S.; Gambini, A.; Renzi, N. y Zizza, R. "Exploiting payments to track Italian economic activity: the experience at Banca d'Italia". Questioni di Economia e Finanza (Occasional Papers), n. 609, Banca d'Italia, Marzo, 2021.

[39] Romero, A.; Arbalejo, N.; Perianes, A.G.; Vicioso, J. y Fernández, S. "La banca digital se abre camino, pero también los peligros y las amenazas". El Mundo, 3 de Mayo, 2025.

[40] Giovine, A.S. "Cybercrime e finanza: una ricerca empirica sul cyber-risk nel mondo bancario". Tesi di Laurea Magistrale, Politecnico di Torino, 2020.

[41] CaixaBank. "Banco Santander, BBVA y CaixaBank se alían para luchar contra el fraude financiero". CaixaBank Newsroom, 21 de Julio, 2023.


COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI