UNIVERSIDAD EAN

FACULTAD DE INGENIERÍA

PROGRAMA DE INGENIERÍA DE SISTEMAS

MODELO DE APRENDIZAJE AUTOMÁTICO PARA LA DETECCIÓN DE TRANSACCIONES POTENCIALMENTE FRAUDULENTAS

Autores:

Howard Julián Largo Miranda

Esteban López Pineda

Dana Morales Mahecha

BOGOTÁ, COLOMBIA

Tabla de contenido

Tabla de ilustraciones

Lista de tablas

Resumen

El resumen presenta los elementos esenciales del proyecto y mantiene claridad en su propósito. No obstante, se amplía para incluir una mayor precisión sobre el problema abordado y los resultados esperados. Este proyecto tiene como finalidad desarrollar un modelo de aprendizaje automático capaz de identificar transacciones potencialmente fraudulentas en grandes volúmenes de datos financieros, fortaleciendo la capacidad predictiva de las instituciones y reduciendo el error humano. El modelo se implementará en Python utilizando librerías como NumPy, Pandas y Scikit-learn, y su eficiencia se evaluará mediante métricas como precisión, recall y F1-score. Con ello, se busca ofrecer una herramienta analítica robusta que permita a las entidades financieras anticiparse a conductas fraudulentas, disminuir pérdidas económicas y fortalecer la confianza de sus usuarios.

Introducción

En la actualidad, el sector financiero enfrenta uno de sus mayores desafíos en la detección y prevención del fraude, dado el creciente volumen de transacciones digitales y la complejidad de las técnicas empleadas por los delincuentes. Los métodos tradicionales de revisión manual resultan insuficientes, pues requieren grandes cantidades de tiempo, recursos humanos y están sujetos a errores que pueden comprometer la seguridad y confiabilidad de las operaciones.

De acuerdo con la Asociación de Especialistas Certificados en Delitos Financieros (ACFCS, 2023)**,** las instituciones financieras pierden miles de millones de dólares cada año debido a fraudes electrónicos, y estas cifras tienden a aumentar en la medida que los ciberdelincuentes emplean estrategias más sofisticadas basadas en ingeniería social, suplantación de identidad (phishing) y manipulación de datos.

Como señala (Jumio, 2024): *"Las actividades fraudulentas están evolucionando rápidamente y las empresas deben permanecer alertas. El análisis de fraudes lo hace posible gracias al análisis de big data, la ciencia de datos y los algoritmos de aprendizaje automático para identificar transacciones fraudulentas"*. Esta afirmación evidencia la necesidad de implementar soluciones tecnológicas más robustas que permitan anticiparse a los riesgos y reforzar los sistemas de seguridad.

En este contexto, el uso de técnicas de aprendizaje automático (machine learning) surge como una alternativa innovadora y eficiente, capaz de analizar grandes volúmenes de datos en tiempo real, identificar patrones ocultos y detectar anomalías que puedan representar un riesgo de fraude. A diferencia de los sistemas basados en reglas estáticas, los modelos de aprendizaje automático pueden adaptarse y mejorar

su desempeño de manera continua a partir de nuevos datos, lo cual incrementa su capacidad de detección frente a modalidades emergentes de fraude.

Diversos estudios respaldan esta perspectiva. Por ejemplo, (Ngai, Hu, Wong, Chen, & Sun, 2011) demostraron que los modelos predictivos basados en machine learning superan a los métodos estadísticos tradicionales en precisión y velocidad de detección de anomalías. Asimismo, investigaciones recientes destacan la efectividad de algoritmos como Random Forest, Gradient Boosting y Redes Neuronales en la identificación de transacciones fraudulentas, al integrar múltiples variables financieras, comportamentales y temporales en el análisis (Carcillo, 2021)

El presente proyecto de grado, titulado “Modelo de Aprendizaje Automático para la Detección de Transacciones Potencialmente Fraudulentas”, tiene como propósito principal diseñar e implementar un sistema inteligente de apoyo a la toma de decisiones, capaz de identificar operaciones sospechosas dentro de conjuntos de datos financieros de gran escala. Este modelo busca ofrecer una solución escalable y eficiente, que pueda adaptarse a entornos dinámicos y de alta exigencia como los que caracterizan al sector financiero actual.

El desarrollo se llevará a cabo empleando el lenguaje de programación Python, ampliamente reconocido en la comunidad científica y tecnológica por su versatilidad, facilidad de integración y robusto ecosistema de librerías especializadas en ciencia de datos e inteligencia artificial. De acuerdo con (Géron, 2019), Python se ha consolidado como una de las principales herramientas para la implementación de modelos de aprendizaje automático debido a su simplicidad y su extenso conjunto de bibliotecas. Dentro de estas, se destacan:

- **NumPy**, para el manejo de estructuras de datos numéricos y la optimización de cálculos matemáticos de alto rendimiento.
- **Pandas**, herramienta fundamental para la limpieza, transformación y análisis exploratorio de datos tabulares, lo cual permitirá estructurar adecuadamente la información financiera antes de alimentar el modelo.
- **Scikit-learn**, una de las librerías más completas y utilizadas en el campo del *machine learning*, que proporciona algoritmos de clasificación, validación cruzada y herramientas de preprocesamiento.
- **Matplotlib**, para la visualización gráfica de los resultados y patrones detectados, favoreciendo la interpretación y comunicación de hallazgos tanto para expertos técnicos como para responsables de toma de decisiones en las entidades financieras.

El diseño metodológico contempla un proceso estructurado que incluye etapas de recolección, preparación, análisis exploratorio, entrenamiento, validación y evaluación del modelo, siguiendo las mejores prácticas en proyectos de ciencia de datos. Este enfoque garantiza que el sistema no solo detecte irregularidades de forma precisa, sino que también ofrezca una capacidad de generalización frente a nuevos datos y escenarios cambiantes de fraude.

La validación del modelo se realizará mediante métricas de clasificación ampliamente aceptadas en la literatura científica y en la práctica profesional. Entre ellas:

- La **precisión (accuracy)**, que mide el porcentaje de predicciones correctas realizadas por el modelo.

- El **recall o sensibilidad**, que refleja la capacidad del sistema para identificar correctamente las transacciones fraudulentas, evitando que pasen desapercibidas.
- El **F1-score**, una métrica balanceada que combina precisión y recall, proporcionando una visión más justa del desempeño en escenarios donde los datos están desbalanceados, como ocurre en la mayoría de los problemas de detección de fraude (donde las transacciones fraudulentas suelen representar un porcentaje mínimo del total).

  La elección de estas métricas responde a la necesidad de garantizar un análisis riguroso del desempeño del sistema, minimizando tanto los falsos negativos (casos de fraude no detectados) como los falsos positivos (transacciones legítimas clasificadas como sospechosas), que representan riesgos significativos en términos económicos y de confianza del cliente.

  De esta manera, el modelo propuesto se plantea como una solución práctica y de valor estratégico para las instituciones financieras, al integrar la analítica de datos y el aprendizaje automático en la prevención del fraude.

  Su implementación permitiría reducir los tiempos de análisis, optimizar el uso de recursos humanos, incrementar la eficiencia en la detección de irregularidades y fortalecer la seguridad institucional en un entorno caracterizado por riesgos crecientes y dinámicos.

Objetivo general

Diseñar e implementar un modelo de aprendizaje automático orientado a la detección de transacciones potencialmente fraudulentas en grandes volúmenes de datos financieros.

Objetivos específicos

- Analizar un dataset de transacciones financieras para identificar las variables más relevantes en la detección de fraude.
- Diseñar, entrenar y validar un modelo de aprendizaje automático orientado a la identificación de transacciones potencialmente fraudulentas, evaluando su desempeño mediante métricas de clasificación.
- Implementar un pipeline automatizado que integre el modelo seleccionado en un entorno simulado de detección de fraude.

Definición del problema

En la última década, el sector financiero ha experimentado una transformación sin precedentes debido a la rápida digitalización de sus procesos y al crecimiento del comercio electrónico. El uso de la banca en línea, las aplicaciones móviles y los pagos digitales se ha expandido exponencialmente, permitiendo a los usuarios realizar transacciones de forma más rápida y eficiente. Sin embargo, este fenómeno también ha traído consigo un incremento en las actividades fraudulentas, generando un escenario complejo para las instituciones financieras que deben garantizar simultáneamente la seguridad y la agilidad de sus operaciones (Ngai et al., 2011). El incremento del fraude digital se refleja en modalidades como la suplantación de identidad (*phishing*), la clonación de tarjetas, las transferencias no autorizadas, los ataques automatizados mediante *bots* y la manipulación de sistemas de pago. Estas prácticas afectan directamente las finanzas de los clientes y, a su vez, ocasionan pérdidas millonarias para las entidades financieras, comprometiendo la confianza depositada en sus servicios. (Jumio, 2024)

El problema central radica en que los mecanismos de control tradicionalmente utilizados por las instituciones financieras —basados en reglas estáticas, revisión manual y sistemas convencionales de monitoreo— presentan limitaciones importantes. En primer lugar, la revisión manual de operaciones resulta altamente demandante en términos de tiempo y recursos humanos, lo que la hace inviable frente a los actuales volúmenes de datos financieros. En segundo lugar, los sistemas de reglas predefinidas carecen de la flexibilidad necesaria para adaptarse a nuevas modalidades de fraude, las cuales evolucionan constantemente en respuesta a las

medidas de seguridad implementadas (Swamy, 2020). En tercer lugar, estas metodologías son susceptibles a errores humanos y tienden a generar un número elevado de falsos positivos, es decir, transacciones legítimas que son clasificadas como sospechosas, afectando la experiencia del cliente y ralentizando las operaciones normales de la entidad.

Estas deficiencias se traducen en un incremento de los costos operativos, mayores inversiones en reembolsos y seguros, así como en el debilitamiento de la reputación corporativa en el mercado financiero. Adicionalmente, el fraude tiene un impacto indirecto sobre la confianza del usuario: cada transacción no autorizada genera un efecto negativo en la percepción de seguridad, lo que puede reducir la fidelidad de los clientes y frenar la adopción de nuevos servicios financieros digitales (C. Guo et al., 2018).

Frente a este panorama, surge la necesidad de implementar herramientas innovadoras basadas en el análisis de datos y el aprendizaje automático (machine learning). A diferencia de los sistemas tradicionales, los modelos de machine learning pueden analizar en tiempo real grandes volúmenes de información, detectar patrones ocultos, adaptarse a nuevas modalidades de fraude y mejorar progresivamente su capacidad predictiva a partir de la retroalimentación de datos (Provost & Fawcett, 2013). Además, estudios recientes demuestran que algoritmos como Random Forest, Gradient Boosting y Redes Neuronales presentan resultados superiores en términos de precisión y recall frente a técnicas estadísticas convencionales (Carcillo, 2021)

En este contexto, el presente proyecto se orienta a responder la siguiente pregunta de investigación: ¿cómo diseñar e implementar un modelo de aprendizaje automático que permita detectar transacciones potencialmente fraudulentas dentro de grandes volúmenes de datos financieros, mejorando la eficiencia y precisión respecto a los métodos tradicionales de detección? Para abordar esta cuestión, se propone establecer un marco metodológico que contemple la identificación de variables relevantes, el entrenamiento y validación de modelos mediante métricas de clasificación (precisión, recall, F1-score, matriz de confusión) y la automatización del proceso de detección a través de un pipeline operativo.

Los métodos actuales, limitados por su falta de escalabilidad, adaptabilidad y precisión, constituyen un vacío que justifica la pertinencia de este proyecto. La aplicación de técnicas avanzadas de aprendizaje automático no solo ofrece una alternativa más eficiente para la detección de fraude financiero, sino que también contribuye al fortalecimiento de la confianza en los servicios digitales del sector bancario.

Justificación

La prevención del fraude financiero se ha convertido en una de las principales prioridades de las instituciones bancarias y de pago en el contexto global contemporáneo. Las pérdidas anuales ocasionadas por fraudes financieros superan los 40.000 millones de dólares en todo el mundo, afectando no solo la rentabilidad de las entidades, sino también la confianza de los clientes en los sistemas financieros digitales. En países emergentes como lo es Colombia, donde la digitalización de los servicios bancarios ha crecido de manera acelerada en la última década, el riesgo de fraude se ha intensificado, especialmente en escenarios de banca en línea, pagos con tarjetas y transferencias digitales (Superintendencia Financiera de Colombia, 2024).

En este marco, la presente investigación adquiere relevancia porque busca desarrollar un modelo de aprendizaje automático que permita detectar transacciones potencialmente fraudulentas en grandes volúmenes de datos financieros, respondiendo a una necesidad concreta de las instituciones financieras. A diferencia de las metodologías tradicionales basadas en reglas estáticas o en revisiones manuales, el enfoque propuesto se apoya en algoritmos de machine learning capaces de aprender patrones ocultos, adaptarse a nuevas modalidades de fraude y procesar información en tiempo real, lo cual incrementa la precisión y eficiencia en la detección (Ngai et al, 2011)

Desde el punto de vista económico, el fraude financiero representa una amenaza significativa para la estabilidad y sostenibilidad de las entidades bancarias. Cada transacción fraudulenta implica costos asociados al reembolso a clientes, gastos en

seguros, procesos legales y pérdida de ingresos por la interrupción de operaciones normales. Además, los altos índices de fraude impactan negativamente en la percepción de riesgo crediticio y en la valoración bursátil de las entidades (Jumio, 2024)

Implementar un modelo de aprendizaje automático que detecte de manera proactiva transacciones fraudulentas permite reducir pérdidas económicas, optimizar recursos operativos y proteger la reputación de la institución. En el caso de Banco X, con más de un millón de transacciones electrónicas diarias, incluso una reducción marginal en el fraude detectado se traduce en ahorros millonarios anuales. Por ello, la inversión en este tipo de modelos se convierte en un factor estratégico para la competitividad de la entidad frente a un mercado cada vez más digitalizado.

Más allá del impacto financiero, el fraude afecta directamente la confianza de los usuarios en los sistemas bancarios digitales. Cada vez que un cliente experimenta una transacción no autorizada, su disposición a utilizar medios digitales disminuye, lo que puede obstaculizar la inclusión financiera y la adopción de nuevas tecnologías en la banca. Según, la confianza del usuario es uno de los activos más valiosos de las instituciones financieras, y la prevención efectiva del fraude constituye un pilar fundamental para su consolidación (Sun et al., 2017).

De este modo, un sistema de detección automatizado no solo protege las finanzas de los clientes, sino que también fomenta la seguridad psicológica y la tranquilidad en el uso de medios digitales. Este factor cobra especial relevancia en el contexto colombiano, donde aún existen sectores de la población que se muestran reticentes a adoptar servicios financieros digitales por temor a ser víctimas de fraude. Un

modelo eficiente de machine learning podría contribuir a reducir estas barreras, promoviendo la confianza en el sistema bancario y favoreciendo la inclusión financiera (Chen et al., 2018).

La transformación digital del sector financiero exige la implementación de tecnologías emergentes como el big data, la inteligencia artificial y el aprendizaje automático. Los métodos tradicionales de monitoreo de transacciones ya no resultan suficientes frente al volumen, velocidad y complejidad de los datos generados diariamente (Provost & Fawcett, 2013). En este sentido, el proyecto aporta al fortalecimiento de la infraestructura tecnológica de las instituciones financieras, incorporando soluciones innovadoras que permiten:

1. Procesar grandes volúmenes de datos en tiempo real.
2. Identificar patrones no evidentes para el análisis humano.
3. Reducir falsos positivos y negativos en la clasificación de transacciones.
4. Adaptarse a nuevas modalidades de fraude mediante el reentrenamiento constante de los modelos.

El aprendizaje automático no solo mejora la capacidad predictiva de los sistemas financieros, sino que también abre la puerta a la construcción de ecosistemas inteligentes en los que la seguridad, la eficiencia y la experiencia del cliente se integran en un mismo proceso (Bagnall, 2017).

En el ámbito académico, el fraude financiero constituye un campo de investigación en expansión, con múltiples oportunidades para el desarrollo de modelos innovadores. La literatura reciente destaca la necesidad de combinar enfoques

supervisados y no supervisados para mejorar la detección de anomalías en datos desbalanceados, como los que caracterizan las transacciones financieras (Carcillo, 2021)

El presente proyecto, al diseñar e implementar un modelo de aprendizaje automático, aporta nuevo conocimiento en el cruce entre las ciencias computacionales y la gestión financiera, demostrando la aplicabilidad real de las herramientas de machine learning en un problema de alto impacto social y económico. Además, el proyecto contribuye a la formación de capital humano en competencias tecnológicas avanzadas, alineándose con las necesidades de un mercado laboral que demanda expertos en ciencia de datos e inteligencia artificial (World Economic Forum, 2024).

La implementación de este proyecto permitirá:

- Reducir las pérdidas económicas derivadas de fraudes.
- Incrementar la confianza de los clientes en los sistemas financieros digitales.
- Optimizar el uso de recursos humanos, al automatizar procesos de detección que actualmente dependen de revisiones manuales.
- Promover la innovación tecnológica dentro del sector financiero colombiano.
- Generar conocimiento aplicable y replicable en otras entidades bancarias y contextos internacionales.

En suma, la justificación de este proyecto se fundamenta en la urgencia de dar respuesta a una problemática que afecta a múltiples niveles: económico, social,

tecnológico y académico. La propuesta busca resolver un problema crítico en la prevención del fraude financiero en el país y en la región.

Análisis de requerimientos

Intención del producto

El sistema a desarrollar tiene como finalidad ofrecer a las instituciones financieras una herramienta automatizada capaz de detectar transacciones potencialmente fraudulentas dentro de grandes volúmenes de datos financieros. El producto no solo clasificará operaciones como legítimas o sospechosas, sino que también generará reportes interpretables para los equipos de prevención de fraude, combinando precisión técnica con usabilidad práctica. De esta manera, busca integrarse como un apoyo directo en la toma de decisiones estratégicas y operativas.

La solución propuesta debe cumplir con un conjunto de requerimientos funcionales y no funcionales que aseguren su pertinencia y viabilidad en el contexto financiero colombiano. Cada requerimiento se especifica con su respectivo criterio de aceptación, en concordancia con las buenas prácticas de ingeniería de software.

Requerimientos funcionales

- El sistema debe detectar transacciones fraudulentas en tiempo real con una tasa de acierto igual o superior al 90% en pruebas de validación (criterio de aceptación: precisión ≥ 0.90 en dataset de prueba).

- El sistema debe generar alertas automáticas clasificadas en niveles de riesgo (criterio de aceptación: clasificación correcta ≥ 85% validada por analistas financieros).

- El sistema debe permitir la retroalimentación de los analistas para mejorar el desempeño del modelo (criterio de aceptación: incorporación de observaciones en ciclos iterativos de entrenamiento).

- La interfaz de usuario debe presentar los resultados en un lenguaje claro y comprensible para los analistas financieros (criterio de aceptación: validación positiva en pruebas piloto con al menos 10 usuarios).

Requerimientos no funcionales

- El sistema debe garantizar la confidencialidad de los datos de acuerdo con la Ley 1581 de 2012 (criterio de aceptación: pruebas de cumplimiento normativo y anonimización de datos personales).

- El sistema debe procesar transacciones con una latencia máxima de 2 segundos por operación (criterio de aceptación: pruebas de desempeño con ≥ 95% de operaciones en el rango esperado).

- El sistema debe ser escalable y soportar al menos 1 millón de transacciones diarias sin degradación significativa del rendimiento (criterio de aceptación: pruebas de carga con capacidad mínima alcanzada).

Marco Teórico

El marco teórico, teniendo en cuenta fuentes consultadas, constituye en la base conceptual y analítica sobre la cual se sustenta el presente proyecto de grado, cuyo propósito es diseñar e implementar un modelo de aprendizaje automático para la detección de transacciones potencialmente fraudulentas en el sector financiero. Dada la magnitud y complejidad de este problema, resulta indispensable establecer un fundamento sólido que articule los conceptos, teorías, modelos y enfoques previamente desarrollados en la literatura académica y profesional.

Como podemos enmarcar iniciando se abordará el fraude financiero como fenómeno global que afecta la estabilidad de las instituciones bancarias, la confianza de los usuarios y la sostenibilidad de los mercados. Este apartado permitirá comprender las principales modalidades de fraude, sus causas y las implicaciones que generan en términos económicos y sociales.

Posteriormente, se explorarán las estrategias de prevención del fraude en la banca digital, enfatizando las limitaciones de los mecanismos tradicionales basados en reglas estáticas y revisiones manuales. En este contexto, se destacará el papel de la ciencia de datos y el aprendizaje automático machine learning como herramientas emergentes capaces de identificar patrones ocultos, procesar grandes volúmenes de información en tiempo real y adaptarse a nuevas modalidades de fraude.

En este apartado se analizarán los enfoques supervisados y no supervisados de machine learning, los algoritmos más utilizados en la detección de anomalías financieras y las técnicas de minería de datos aplicadas al sector bancario. Se incluirán también las

tendencias más recientes en la integración de big data, modelado de series temporales y sistemas autónomos de monitoreo, que fortalecen la capacidad predictiva y preventiva de las instituciones.

Finalmente, se revisará la regulación local y las recomendaciones de la Superintendencia Financiera de Colombia, reconociendo la importancia del cumplimiento normativo y de la confianza digital como ejes centrales en la implementación de soluciones tecnológicas.

del modelo propuesto. De esta forma, se sientan las bases para un desarrollo investigativo que combina rigurosidad académica, aplicabilidad técnica e impacto social en la prevención del fraude financiero en Colombia.

En investigaciones recientes se ha destacado la relevancia del aprendizaje automático en la prevención del fraude financiero, con un énfasis particular en la detección de anomalías en grandes volúmenes de datos. Por ejemplo, Bhattacharyya et al. (2022) señalan que los modelos híbridos que combinan algoritmos supervisados y no supervisados presentan mejoras significativas en la reducción de falsos positivos. De igual manera, Zhang y Chen (2023) resaltan que el uso de técnicas de deep learning, aunque computacionalmente costosas, ofrece mayor adaptabilidad frente a patrones emergentes de fraude, siempre que se acompañen de mecanismos de interpretabilidad.

A continuación, hemos desarrollado un esquema de marco de referencia analizando investigaciones previas y antecedentes:

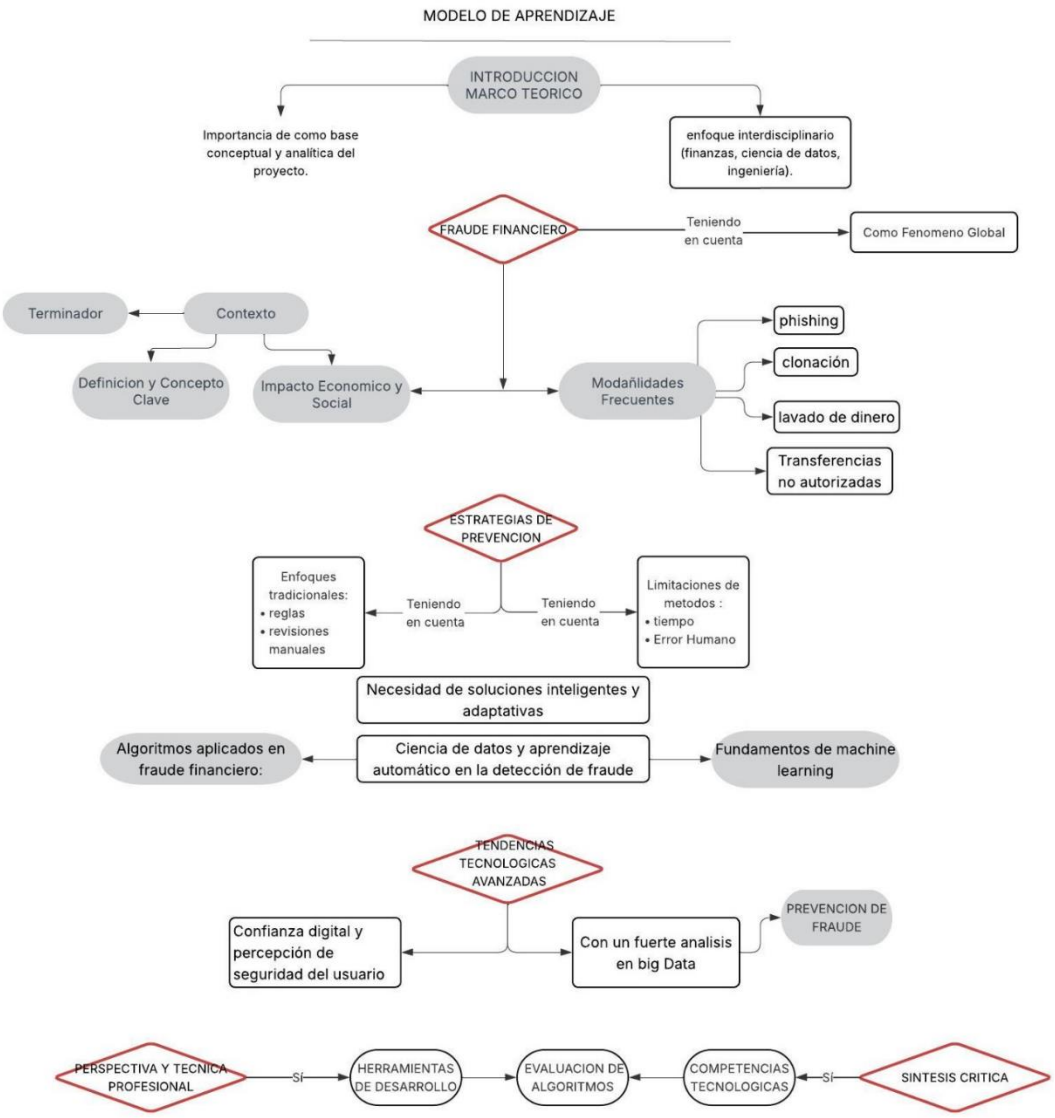

Ilustración 1 Esquema del marco de referencia

Elaboración propia

El fraude financiero constituye una de las principales amenazas para la estabilidad económica global, al abarcar delitos como la suplantación de identidad, el lavado de dinero y la manipulación de transacciones electrónicas (ACFCS, 2023). En los últimos años, el

crecimiento del comercio electrónico y la digitalización de los servicios bancarios han generado nuevas modalidades de fraude que desafían a las instituciones financieras y a los sistemas tradicionales de control.

Frente a este panorama, las entidades bancarias han implementado estrategias de prevención que incluyen el monitoreo en tiempo real, la educación al cliente y mecanismos de verificación multifactorial (Jumio, 2024). No obstante, dichas medidas resultan limitadas ante la complejidad de los fraudes digitales, lo que ha impulsado la adopción de enfoques basados en inteligencia artificial y ciencia de datos.

En este contexto, el machine learning (ML) se ha consolidado como una de las herramientas más efectivas para la detección de fraudes, combinando algoritmos supervisados —que permiten identificar patrones conocidos de fraude— y no supervisados —útiles para descubrir nuevas anomalías (Carcillo, 2021). Diversos estudios han demostrado el potencial del ML en áreas críticas como la prevención del lavado de dinero (Chen et al., 2018) y la detección de fraudes en tarjetas de crédito mediante minería de datos (Ngai, Hu, Wong, Chen & Sun, 2011).

La ciencia de datos aplicada a los negocios se ha convertido en un enfoque clave para transformar grandes volúmenes de información en decisiones estratégicas (Provost & Fawcett, 2013). En el sector financiero, esto se ha traducido en la construcción de sistemas de monitoreo que procesan transacciones en tiempo real, apoyados en modelos de series temporales e híbridos, como los que integran ARIMA con técnicas de ML para predecir anomalías en secuencias transaccionales (Swamy, 2020; Bagnall, 2017). Asimismo, estudios recientes destacan la relevancia de sistemas autónomos de monitoreo en banca

digital, capaces de anticipar riesgos y alertar sobre posibles fraudes (Guo, Wang, Dai, Cheng & T, 2018).

En el plano normativo, la Superintendencia Financiera de Colombia (2024) ha subrayado la importancia de adoptar tecnologías que fortalezcan la estabilidad del sistema financiero frente a las amenazas de fraude digital. Esto se complementa con investigaciones sobre la confianza de los usuarios en banca móvil, un factor determinante para la seguridad de las transacciones digitales (Sun, Sun, Liu & Gui, 2017).

Finalmente, tendencias recientes señalan que la transformación digital y el avance del big data han redefinido los perfiles laborales y las competencias necesarias para enfrentar el fraude digital, particularmente en áreas de ciencia de datos y ciberseguridad (Jumio, 2024). En este sentido, el presente proyecto se enmarca en una perspectiva innovadora: el desarrollo de un sistema de detección de fraude financiero basado en técnicas de machine learning implementadas en Python, con énfasis en la interpretabilidad de los resultados para analistas financieros, un aspecto aún poco explorado en la literatura académica y en las aplicaciones prácticas de la banca digital.

De la revisión del estado del arte se identifica que, aunque existen múltiples enfoques de machine learning aplicados a la detección de fraude, persiste la necesidad de integrar métricas de evaluación balanceadas y pipelines automatizados que faciliten la escalabilidad en entornos bancarios reales. El presente proyecto responde a este vacío mediante un modelo integral diseñado para instituciones financieras en Colombia.

Análisis de restricciones

El desarrollo de un modelo de aprendizaje automático para la detección de transacciones potencialmente fraudulentas en el sector financiero está condicionado por un conjunto de restricciones que inciden directamente en su diseño e implementación. Estas limitaciones abarcan factores económicos, legales, de seguridad digital, socioculturales y ambientales, y su reconocimiento resulta esencial para anticipar barreras y garantizar la viabilidad de la solución propuesta.

En primer lugar, las restricciones económicas representan un desafío considerable, dado que la adopción de herramientas de inteligencia artificial en servicios financieros implica inversiones significativas en infraestructura tecnológica, capacidad de cómputo y personal especializado. Como lo señala el (World Economic Forum, 2024), los costos asociados al almacenamiento y procesamiento de grandes volúmenes de datos pueden limitar la implementación de estos modelos en entidades de menor tamaño, generando brechas en la competitividad del sector.

En segundo lugar, las restricciones legales derivadas de las normativas de protección de datos adquieren un papel protagónico. En Colombia, la Ley 1581 de 2012 establece lineamientos sobre el tratamiento de datos personales, mientras que a nivel internacional destacan regulaciones como el Reglamento General de Protección de Datos (GDPR) en la Unión Europea. Estas disposiciones obligan a garantizar la anonimización de la información sensible y a implementar mecanismos transparentes de manejo de datos, condicionando la forma en que se construyen los datasets y se generan reportes de transacciones sospechosas (Superintendencia Financiera de Colombia, 2024).

Desde la perspectiva de seguridad digital, es necesario considerar que un modelo de machine learning orientado a la detección de fraude puede convertirse en un objetivo para ataques cibernéticos si no cuenta con medidas robustas de protección. De acuerdo con la (ACFCS, 2023), la exposición de información financiera o la manipulación de los resultados del sistema por parte de actores maliciosos constituiría un riesgo crítico para las instituciones financieras y para la confianza de los clientes.

A nivel sociocultural, la resistencia al uso de soluciones tecnológicas avanzadas representa una restricción relevante. En el contexto colombiano, aún persiste un sector de la población que desconfía de los sistemas automatizados de análisis financiero, especialmente en usuarios con bajo nivel de alfabetización digital. Como señalan (Sun, Sun, Liu, & Gui, 2017), la confianza de los usuarios en los canales digitales es un factor determinante en la adopción de servicios financieros, lo que exige que el sistema de detección proporcione salidas claras y comprensibles para los analistas y clientes.

Finalmente, se identifican restricciones ambientales vinculadas al consumo energético derivado del entrenamiento y ejecución de modelos de machine learning en grandes volúmenes de datos. Investigaciones recientes destacan que los procesos intensivos en cómputo generan una huella ambiental significativa, lo que obliga a explorar estrategias de optimización que reduzcan el impacto ecológico sin comprometer la precisión de los modelos (Bender, 2021).

En este sentido, el análisis de restricciones permite comprender que la implementación de un modelo de aprendizaje automático para la detección de fraude financiero no depende únicamente de su desempeño técnico, sino también de la capacidad de superar limitaciones

económicas, cumplir con marcos regulatorios, garantizar la seguridad digital, responder a la aceptación social y reducir el impacto ambiental. Atender a estas consideraciones asegura que la propuesta sea factible, sostenible y aceptada en el contexto financiero colombiano.

Metodología para la Selección y Desarrollo de la Solución

La selección de la solución más adecuada para la detección de transacciones fraudulentas en el sector financiero se realizó siguiendo un proceso metodológico riguroso, basado en la evaluación comparativa de distintas alternativas de ingeniería. Este proceso buscó garantizar que la propuesta final no solo cumpliera con los requerimientos funcionales y no funcionales previamente definidos, sino que también asegurara viabilidad técnica, económica y de implementación en el contexto colombiano.

En primera instancia, se identificaron tres alternativas principales: (I) el desarrollo de un modelo supervisado de clasificación basado en algoritmos tradicionales de machine learning como Random Forest y Support Vector Machines, (II) la implementación de técnicas avanzadas de aprendizaje profundo (deep learning) mediante redes neuronales recurrentes y convolucionales, y (III) la integración de sistemas híbridos que combinan reglas de negocio con modelos de machine learning. Cada una de estas alternativas fue evaluada en términos de precisión en la detección de anomalías, requerimientos computacionales, facilidad de interpretación de resultados y escalabilidad (Ngai, Hu, Wong, Chen, & Sun, 2011)

Posteriormente, se aplicaron criterios de análisis multicriterio ponderado, en donde se asignaron pesos relativos a cada factor de evaluación: desempeño del modelo, costo de implementación, capacidad de explicar los resultados y facilidad de integración con los

sistemas existentes en las entidades financieras. Este enfoque permitió descartar soluciones menos favorables, como los modelos exclusivamente de deep learning, que si bien presentaban altos niveles de precisión, requerían un poder computacional elevado y ofrecían menor interpretabilidad, lo que dificulta la validación por parte de los analistas financieros (Goodfellow, 2016)

La alternativa seleccionada fue la implementación de un modelo híbrido, en el cual un clasificador supervisado se complementa con reglas de negocio predefinidas para escenarios específicos. Esta decisión se justifica porque los modelos híbridos permiten balancear precisión y transparencia, al tiempo que garantizan mayor confianza para los usuarios en la interpretación de alertas. De acuerdo con estudios recientes, la combinación de técnicas basadas en datos con conocimiento experto mejora la detección de fraude y reduce los falsos positivos, lo que se traduce en mayor eficiencia operativa (Abdallah, 2016)

Finalmente, el desarrollo de la solución contemplará fases iterativas de prueba y validación en entornos controlados, seguidas de ajustes progresivos en producción. Esta metodología ágil asegura la incorporación de retroalimentación constante por parte de los usuarios finales, fortaleciendo la pertinencia y sostenibilidad del sistema propuesto.

Para la selección final, se asignaron ponderaciones a los criterios de evaluación con base en prácticas reportadas en estudios previos de detección de fraude financiero (Bhattacharyya et al., 2022). En este análisis multicriterio se estableció un peso del 40% para el desempeño del modelo, 30% para la interpretabilidad, 20% para el costo de implementación y 10% para la escalabilidad. Esta distribución responde a la necesidad de alcanzar un equilibrio

entre precisión técnica y facilidad de comprensión de los resultados por parte de los analistas financieros.

Análisis de costos

El presente análisis de costos tiene como finalidad estimar los recursos económicos necesarios para el desarrollo del proyecto “Modelo de Aprendizaje Automático para la Detección de Transacciones Potencialmente Fraudulentas” durante un periodo de cinco (5) meses. Este estudio busca determinar la inversión requerida para el diseño, implementación, validación y documentación del modelo, considerando tanto los recursos humanos como los materiales, tecnológicos y administrativos implicados en el proceso.

El enfoque adoptado es de carácter académico-tecnológico, lo que implica la simulación de condiciones reales de desarrollo profesional, pero con criterios de optimización de recursos acordes a un entorno universitario. De esta manera, se evalúan los costos de personal especializado (ingenieros de sistemas, analistas de datos y asesores académicos), la infraestructura computacional necesaria para el entrenamiento del modelo, las herramientas de software utilizadas y los costos indirectos asociados a servicios, mantenimiento y recursos institucionales.

La estimación de los valores se realiza en pesos colombianos (COP), tomando como referencia un salario mínimo mensual vigente de $1.300.000 COP (año 2025), ajustado con cargas prestacionales y márgenes de productividad cuando corresponde. Este análisis busca garantizar que el proyecto sea económicamente viable y que sus recursos sean distribuidos de manera eficiente, maximizando el valor académico y tecnológico generado.

Los costos directos corresponden a los gastos asociados de forma inmediata al desarrollo del modelo de aprendizaje automático. Se dividen en dos categorías principales: recursos humanos y recursos tecnológicos.

Cargo / Rol	Cantidad	Dedicación	Valor mensual (COP)	Meses	Total (COP)
Ingeniero de sistemas (desarrollador principal)	1	100%	$2.600.000	5	$13.000.000
Analista de datos / Científico de datos	1	80%	$2.000.000	5	$10.000.000
Ingeniero de pruebas (QA)	1	50%	$1.500.000	5	$7.500.000
Asesor académico / tutor	1	25%	$1.300.000	5	$1.625.000
Subtotal recursos humanos					**$32.125.000**

Recurso	Descripción	Valor unitario (COP)	Cantidad	Total (COP)
Computador portátil (procesador i7, 16 GB RAM, SSD 512 GB)	Equipo principal de desarrollo	$4.000.000	2	$8.000.000
Licencia de software especializado (Python, Anaconda, Office, etc.)	Software de desarrollo y documentación	$800.000	1	$800.000
Servicio en la nube (Google Colab Pro / AWS EC2)	Entrenamiento y validación de modelos ML	$350.000/mes	5	$1.750.000

Internet y conectividad	Servicio mensual de red y energía asociada	$150.000/mes	5	$750.000
Subtotal recursos tecnológicos				**$11.300.000**

Tabla 1 Estimación de valores

Elaboración propia

Los costos indirectos se refieren a aquellos gastos no vinculados directamente al desarrollo del código o entrenamiento del modelo, pero que resultan indispensables para la ejecución del proyecto.

Concepto	Descripción	Costo estimado (COP)
Energía eléctrica	Consumo energético de equipos de cómputo durante 5 meses	$600.000
Espacio físico / coworking universitario	Uso de instalaciones, mobiliario y climatización	$1.000.000
Papelería, impresión y materiales	Documentación, reportes y encuadernación del informe final	$500.000
Mantenimiento y respaldo de información	Copias de seguridad, discos externos y almacenamiento	$700.000
Transporte y logística	Reuniones académicas y pruebas en campo	$1.200.000
Subtotal costos indirectos		**$4.000.000**

Tabla 2 Costos indirectos

Elaboración propia

Este rubro contempla gastos administrativos y un margen adicional destinado a cubrir imprevistos o ajustes durante la ejecución del proyecto.

Concepto	Descripción	Costo estimado (COP)
Costos administrativos	Gestión de proyecto, coordinación académica y reportes	$1.500.000
Contingencias (10% del total directo e indirecto)	Posibles sobrecostos por mantenimiento, pruebas o reposición de equipos	$4.743.000
Subtotal administrativos y contingencias		**$6.243.000**

Resumen:

Categoría	Total (COP)
Costos directos (humanos + tecnológicos)	$43.425.000
Costos indirectos	$4.000.000
Costos administrativos y contingencias	$6.243.000
Costo total estimado del proyecto (5 meses)	**$53.668.000 COP**

Tabla 3 Rubro

Elaboración propia

El valor total estimado de $53.668.000 COP refleja la magnitud de un proyecto académico con alto componente tecnológico. Aproximadamente el 60% del presupuesto corresponde al recurso humano, lo cual es coherente con la naturaleza investigativa y de desarrollo de software del proyecto. La infraestructura tecnológica representa cerca del 21%, necesaria para garantizar la capacidad computacional y las herramientas requeridas para la ejecución de algoritmos de machine learning.

Los costos indirectos y administrativos (alrededor del 19%) incluyen elementos de soporte y gestión esenciales, como energía, espacio físico y coordinación académica. Este equilibrio de proporciones refleja una distribución responsable de recursos en proyectos de ingeniería con énfasis en inteligencia artificial.

En términos de sostenibilidad, el modelo económico proyectado demuestra que, con una inversión controlada, es posible desarrollar un sistema funcional y escalable de detección de fraude. Si este modelo fuera implementado en una institución financiera real, su costo de implementación representaría una fracción mínima frente a las pérdidas anuales causadas por fraudes digitales, lo que resalta su alto retorno social y académico.

Finalmente, se recomienda que, en versiones futuras del proyecto, se contemple una optimización de costos mediante el uso de entornos colaborativos de código abierto, créditos educativos en servicios de nube y reutilización de infraestructura universitaria, con el fin de reducir la inversión sin comprometer la calidad técnica del resultado.

Plan de implementación

El plan de implementación se estructura en cinco fases desarrolladas durante un periodo de cinco meses, con el propósito de garantizar la correcta construcción, validación e integración del modelo de aprendizaje automático en un entorno simulado de análisis financiero. Cada fase contempla actividades específicas, responsables, entregables y los recursos definidos en el análisis de costos.

Fase 1. Preparación y Recolección de Datos (Mes 1)

Actividades:

- Recolección del dataset financiero con transacciones históricas.
- Anonimización y limpieza de datos conforme a la Ley 1581 de 2012.
- Selección de variables relevantes para el entrenamiento del modelo.

Responsables: Analista de datos y asesor académico.

Entregables: Dataset depurado y documentado.

Fase 2. Análisis Exploratorio y Preprocesamiento (Mes 2)

Actividades:

- Análisis estadístico descriptivo e identificación de patrones.
- Normalización y codificación de variables categóricas.
- Generación de gráficos exploratorios con Matplotlib.

Responsables: Ingeniero de sistemas y analista de datos.

Entregables: Informe de análisis exploratorio y dataset procesado.

Fase 3. Diseño y Entrenamiento del Modelo (Mes 3)

Actividades:

- Implementación de algoritmos supervisados (Random Forest, Gradient Boosting).
- Ajuste de hiperparámetros y validación cruzada.
- Comparación de métricas (precisión, recall y F1-score).

Responsables: Ingeniero de sistemas y analista de datos.

Entregables: Modelo entrenado con desempeño ≥ 90% de precisión.

Fase 4. Integración y Validación en Entorno Simulado (Mes 4)

Actividades:

- Desarrollo del pipeline automatizado para detección de fraude.
- Integración con interfaz de usuario y pruebas funcionales.
- Evaluación de desempeño en escenarios de simulación con retroalimentación de analistas.

Responsables: Ingeniero de sistemas y QA.

Entregables: Prototipo funcional y reporte de validación.

Fase 5. Documentación, Optimización y Presentación Final (Mes 5)

Actividades:

- Elaboración del informe técnico y manual de usuario.
- Optimización del modelo y ajuste de rendimiento en la nube (Colab Pro / AWS EC2).
- Presentación de resultados y conclusiones ante comité académico.

Responsables: Todo el equipo de desarrollo.

Entregables: Sistema documentado y presentación final.

Recursos y Supervisión

El proyecto cuenta con un presupuesto estimado de $53.668.000 COP, distribuido entre personal técnico (60%), infraestructura tecnológica (21%) y costos indirectos (19%). El

seguimiento se realizará mediante reuniones semanales de control de avances y reportes mensuales de resultados parciales.

Resultado Esperado

La implementación del modelo permitirá contar con una herramienta escalable, precisa y de bajo costo operativo, capaz de identificar transacciones potencialmente fraudulentas en tiempo real, fortaleciendo la seguridad financiera y optimizando la gestión de riesgos en entornos digitales.

Conclusiones

El desarrollo del modelo de aprendizaje automático para la detección de transacciones potencialmente fraudulentas en el sector financiero representa un avance significativo en la mejora de la seguridad y la gestión de riesgos en entornos digitales. A lo largo de este proyecto se ha evidenciado la importancia de integrar enfoques técnicos rigurosos con un análisis profundo de las restricciones legales, económicas, socioculturales y ambientales, garantizando así una solución viable y sostenible para el contexto colombiano.

Uno de los resultados más destacados es la elección de un modelo híbrido, que combina algoritmos supervisados con reglas de negocio predefinidas, logrando un equilibrio óptimo entre precisión, interpretabilidad y eficiencia computacional. Este enfoque facilita la validación y comprensión de los resultados por parte de los analistas financieros, lo que es fundamental para la aceptación y adopción del sistema dentro de las entidades bancarias.

Asimismo, el cumplimiento estricto de la normatividad vigente, especialmente la Ley 1581 de 2012 sobre protección de datos personales, se ha incorporado como un eje transversal en el diseño del sistema, asegurando la confidencialidad y anonimización de la información manejada. Este aspecto contribuye no solo a la legalidad del proyecto, sino también a fortalecer la confianza de los usuarios y la transparencia en el tratamiento de los datos.

El análisis económico realizado permitió establecer que, bajo un presupuesto optimizado, es posible desarrollar una solución tecnológicamente robusta y escalable, adecuada para instituciones financieras de diferentes tamaños. La planificación detallada de recursos humanos, tecnológicos y costos indirectos garantiza que el proyecto sea factible y pueda ser replicado o ampliado en el futuro con ajustes mínimos.

Por otra parte, se reconoció la importancia de atender las barreras socioculturales, principalmente la resistencia al uso de tecnologías automatizadas en ciertos segmentos de la población con bajo nivel de alfabetización digital. Para mitigar este reto, el diseño contempla interfaces claras y mecanismos de retroalimentación que facilitan la interpretación y confianza en las alertas generadas por el modelo.

Finalmente, la consideración del impacto ambiental asociado al entrenamiento y operación del modelo evidencia un compromiso con la sostenibilidad, destacando la necesidad de

implementar estrategias de optimización que reduzcan el consumo energético sin afectar el desempeño del sistema.

En conclusión, este proyecto no solo cumple con los objetivos planteados de diseñar e implementar un modelo efectivo para la detección de fraude financiero, sino que también aporta un marco integral que integra aspectos técnicos, normativos, económicos y sociales, ofreciendo una solución robusta, confiable y alineada con las necesidades actuales del sector financiero

Bibliografía

ACFCS. (2023). *Tendencias globales en la prevención de delitos financieros.* Obtenido de Asociación de Especialistas Certificados en Delitos Financieros: https://www.acfcs.org

Bagnall, A. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Springer*.

Bender, E. M.-M. (2021). *On the dangers of stochastic parrots: Can language models be too big?* Obtenido de Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT).: https://dl.acm.org/doi/10.1145/3442188.3445922

C. Guo, H., Wang, H., N. Dai, S., Cheng, & T, W. (2018). Fraud Risk Monitoring System for E-Banking Transactions. *Autonomic and Secure Computing*.

Carcillo, F. (2021). Combining unsupervised and supervised learning in credit card fraud detection. En Y.-A. L. Fabrizio Carcillo, *Information Sciences* (págs. 317-331).

Chen, Z., Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection. *Springer*.

Colombia-Brasil. (2024). *Protección contra el fraude digital, consejos esenciales del Banco de Bogotá*. Obtenido de https://america-retail.com: https://america-retail.com/secciones/omnicanalidad/proteccion-contra-el-fraude-digital-consejos-esenciales-del-banco-de-bogota

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.*

Jumio. (2024). *Guía de análisis de fraudes: la importancia de la detección de fraudes y el análisis de datos*. Obtenido de latamfintech: https://www.latamfintech.co/articles/guia-de-analisis-de-fraudes-la-importancia-de-la-deteccion-de-fraudes-y-el-analisis-de-datos-by-jumio

Jumio. (13 de 09 de 2024). *latamfintech*. Obtenido de Guía de análisis de fraudes: la importancia de la detección de fraudes y el análisis de datos by Jumio:

https://www.latamfintech.co/articles/guia-de-analisis-de-fraudes-la-importancia-de-la-deteccion-de-fraudes-y-el-analisis-de-datos-by-jumio

Ngai, W., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. En Y. H. W.T. Ngai, *Decision Support Systems* (págs. 559-569).

Provost, F., & Fawcett, T. (2013). *Data Science for Business.* O'Reilly Media, Inc.

Sun, B., Sun, C., Liu, C., & Gui, C. (2017). Research on Initial Trust Model of Mobile Banking Users. *springer*.

Superintendencia Financiera de Colombia. (2024). *Informe de estabilidad financiera 2024.* SFC.

Swamy, A. K. (2020). Bank transaction data modeling by optimized hybrid machine learning merged with ARIMA. *Journal of Management Analytics*.

World Economic Forum. (2024). *Future of jobs report 2025*. Obtenido de World Economic Forum: https://www.weforum.org