UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE GUAYAQUIL

CARRERA DE COMPUTACIÓN

Propuesta de un modelo para análisis de prevención de lavado de dinero en Ecuador basado en Machine Learning

Trabajo de titulación previo a la obtención del
Título de Ingeniero en Ciencias de la Computación

AUTOR: GEOVANNY DAVID BAJAÑA MORENO

TUTOR: JOE FRAND LLERENA IZQUIERDO, ING., MSC.

Guayaquil – Ecuador

2024

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Geovanny David Bajaña Moreno con documento de identificación N° 0929255065 manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Guayaquil, 31 de enero del año 2024

Atentamente,

Geovanny David Bajaña Moreno
0929255065

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Yo, Geovanny David Bajaña Moreno con documento de identificación N° 0929255065, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Artículo Académico: "Propuesta de un modelo para análisis de prevención de lavado de dinero en Ecuador basado en Machine Learning", el cual ha sido desarrollado para optar por el título de: Ingeniero en Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Guayaquil, 31 de enero del año 2024

Atentamente,

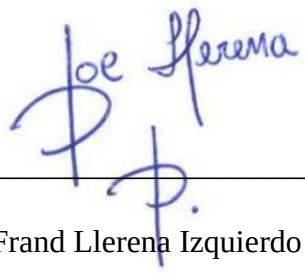
Geovanny David Bajaña Moreno
0929255065

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Joe Frand Llerena Izquierdo con documento de identificación N° 0914884879, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: Propuesta de un modelo para análisis de prevención de lavado de dinero en Ecuador basado en Machine Learning, realizado por Geovanny David Bajaña Moreno con documento de identificación N° 0929255065, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Guayaquil, 31 de enero del año 2024

Atentamente,



Joe Frand Llerena Izquierdo
0914884879

DEDICATORIA

Dedico este trabajo a Dios por permitirme darme fuerzas para estudiar y trabajar y a la vez guiarme en un camino hacia el éxito, Él es mi motor para seguir en mis conocimientos y realizar este trabajo con mucho amor, paciencia y alegría.

En segundo lugar dedico este trabajo de titulación a mi abuela que fue un impulso fundamental en mi desarrollo profesional. No fue fácil todo este proceso de estudio ya que hubo muchas adversidades en todos estos años de estudio y me siento muy orgulloso de poder culminar este proceso con alegría y entusiasmo.

Dedico este articulo a las personas que estuvieron presentes en esta etapa de mi vida ya que fue muy importante en este trabajo de estudio para mi desarrollo personal y para mí es un logro más, que las circunstancia me han llevado a culminar por consejos de personas muy queridas a lo largo de estos años.

Geovanny David Bajaña Moreno

AGRADECIMIENTO

Agradezco a Dios, hermanos, padres, familiares, docentes y amigos por todo el apoyo incondicional que me han brindado durante esta maravillosa etapa de mi vida, y los conocimientos aprendidos de cada persona que he conocido.

A si mismo quiero agradecer a la institución donde trabajo por haberme permitido estudiar está carrera, donde siempre la tecnología evoluciona cada día más. Todas las personas que trabajan conmigo siempre me aconsejan y aprendo de ellos que tienen más experiencia.

Agradezco por todas las experiencias vividas, alegrías, aprendizajes y más. Esto me hizo entender que cada persona es capaz de realizar lo que se proponga en su vida, sobre encima de los obstáculos, pero con amor y dedicación podemos culminar nuestros objetivos, nunca hay que desmayar siempre persiguiendo nuestros sueños. Estoy muy contento de haber realizado este trabajo con esfuerzo y disciplina.

¡Muchas gracias por todo!

Geovanny David Bajaña Moreno

RESUMEN

El objetivo general es elaborar un modelo en prevención de lavado de dinero para identificación de patrones y mejorar la toma de decisiones basado en tecnología Machine Learning. La metodología para la extracción y análisis de artículos científicos, se exploran bibliotecas digitales, además se adopta el algoritmo Random Forest de acuerdo a la extracción y análisis de artículos científicos que recomiendan el algoritmo más apropiado. Se obtuvo 22 artículos que sirven para el análisis, y se halló contra el lavado de dinero: El mayor sector que utiliza ML es la Banca, el primer algoritmo ML que utilizan es Random Forest, las mayores entidades que utilizan ML son Comerciales, la primera actividad que se realiza es el Fraude, la primera técnica que se realiza es la Clasificación, la mayor herramienta utilizada con algoritmos ML es Deep Learning. Se propone un modelo basado en el clasificador Random Forest porque se basa en la puntuación de probabilidad, la selección de los vectores más extraños, y la selección de las transacciones agregadas menos extrañas. Por supuesto, que existe mucho margen de mejora en el modelo que se propone. Además, se necesita un conjunto de datos muy grande con campos o atributos profundos.

Palabras claves: Inteligencia Artificial, Aprendizaje Automático, Modelo contra lavado de dinero, transacciones extrañas.

ABSTRACT

The general objective is to develop a model in the prevention of money laundering to identify patterns and improve decision-making based on Machine Learning technology. The methodology for the extraction and analysis of scientific articles, digital libraries are explored, and the Random Forest algorithm is adopted according to the extraction and analysis of scientific articles that recommend the most appropriate algorithm. A total of 22 articles were obtained for analysis, and the following anti-money laundering items were found: The largest sector that uses ML is Banking, the first ML algorithm they use is Random Forest, the largest entities that use ML are Commercial, the first activity that is carried out is Fraud, the first technique that is performed is Classification, the largest tool used with ML algorithms is Deep Learning. A model based on the Random Forest classifier is proposed because it is based on probability scoring, selection of the strangest vectors, and selection of the least bizarre aggregate transactions. Of course, there is a lot of room for improvement in the proposed model. In addition, you need a very large dataset with deep fields or attributes.

Key words: Artificial Intelligence, Machine Learning, Anti-Money Laundering Model, Strange Transactions.

ÍNDICE DE CONTENIDO

1. INTRODUCCIÓN 10
2. REVISIÓN DE LITERATURA 13
2.1. Machine Learning (ML) 13
2.2. Algoritmos de Machine Learning 13
2.3. Lavado de dinero (LD) 13
2.4. Casos de prevención del lavado de dinero mediante ML 14
3. METODOLOGÍA 15
4. RESULTADOS 17
5. DISCUSIÓN 28
6. CONCLUSIÓN 29
REFERENCIAS 30

1. INTRODUCCIÓN

El lavado de dinero (LD) es cualquier actividad que mantiene los ingresos de actividades ilícitas como tráfico de drogas, evasión fiscal o tráfico de armas; y se realizan los depósitos en el sistema financiero de cualquier país, mediante varias operaciones que disimulan sus orígenes ilegítimos; el lavado de dinero perturba a cualquier economía en el mundo y mueve flujos financieros ilegales que están entre 2,1% y 4% del Producto Bruto Mundial. El seguimiento de las transacciones por lavado de dinero es un conjunto de actividades ejecutadas por analistas y sistemas informáticos para inspeccionar los movimientos de los clientes para detectar conductas sospechosas relacionadas al blanqueo de dinero (Pérez González, 2021; Toala Indio, 2021); los movimientos financieros pueden ser transacciones con tarjeta de crédito, transferencias bancarias, transacciones en banca de inversión, compra de acciones, compra de bienes, establecer empresas y otros derivados (Rosero Tejada, 2021). Existen sistemas informáticos para combatir el LD basados en parámetros o reglas que monitorean los comportamientos inusuales; el sistema informático genera alertas al momento de sobrepasar ciertos límites, por ejemplo, en caso de que la transacción de dinero es mayor que cierta cantidad entonces se genera una alerta. Los analistas revisan la información en forma más sencilla; aunque las técnicas para el lavado de dinero y otros delitos están en continua evolución, entonces las reglas en los sistemas informáticos también deben actualizarse; además también deben detectar comportamientos anormales o desconocidos (Labanca et al., 2022; Tacuri López, 2021; Vera Navas, 2021; Yaya et al., 2021) La lucha contra el LD reduce esa conversión de los lavadores de dinero en activos legales mediante leyes, técnicas, tecnología y regulaciones; las instituciones financieras y organizaciones realizan el seguimiento de las transacciones y aplican el seguimiento de patrones o transacciones anormales de los clientes; las aplicaciones informáticas emiten las alertas por eventos como tamaño de transacción, ubicación de la transacción, variación en la conducta del cliente, riqueza sub-realista, cantidad de bienes, cantidad de negocios. Existe una tasa de transacciones en falsos positivos que superan el 95%; esto resulta costoso y consume mucho tiempo, la cantidad de transacciones están en aumento por las aplicaciones en línea, y las técnicas basadas en reglas dificultan el seguimiento (Oztas et al., 2022), (Yahaya et al., 2020).

Múltiples organizaciones de gobiernos extranjeros y organizaciones privadas confirman que Machine Learning (ML) es muy satisfactorio para optimizar las actividades contra el LD; ML tiene como objetivo principal el identificar nuevos tipos de LD y optimizar el uso de recursos

contra el LD. La ley internacional contra el LD se basa en recomendaciones, en este lineamiento cualquier acción con productos delictivos es un LD, además, las empresas deben conocer al cliente, el riesgo de LD relacionado e informar conductas sospechosas (Jensen & Iosifidis, 2023). Los enfoques ML mantienen la reputación de las empresas, minimizan los costos operativos, reducen los falsos positivos en transacciones, emite resultados eficientes y compensan los requisitos de las entidades reguladoras; al utilizar algoritmos que identifican el LD, se depende de las características que tienen los datos transaccionales que se traduce en efectividad del método. El método supervisado en ML es más sencillo de evaluar por las etiquetas que tiene el conjunto de datos y deben ser etiquetas de alta calidad; el método no supervisado se adapta a las variaciones en la conducta de los clientes y detecta actividades sospechosas (Oztas et al., 2022; Zia et al., 2019)

En Ecuador, la "Ley Prevención de Lavado de Activos y del Financiamiento de Delitos" en su artículo 5 especifica que las instituciones están obligados a entregar reportes de las transacciones de dinero, además el artículo 8 especifica que las personas que salgan o entren a Ecuador con cantidad superior a diez mil dólares americanos (o equivalente) deben declararlo a las autoridades aduaneras (Asamblea-Ecuador, 2023).

El objetivo general es elaborar un modelo en prevención de lavado de dinero para identificación de patrones y mejorar la toma de decisiones basado en tecnología Machine Learning.

Los objetivos específicos son:

- Extraer artículos científicos sobre Machine Learning y Lavado de dinero mediante la exploración en bibliotecas digitales de los últimos cinco años, para conocer el uso de la tecnología en un área específica.
- Analizar los resultados sobre la tecnología Machine Learning en la prevención del lavado de dinero mediante la deducción e inducción de los análisis de artículos científicos, para determinar el algoritmo ML más apropiado a utilizar.
- Elaborar un modelo teórico en análisis de prevención de lavado de dinero en Ecuador basado en un algoritmo de Machine Learning, para identificación de patrones y mejorar la toma de decisiones.

El ML optimiza las técnicas contra el lavado de dinero; un modelo de Machine Learning extrae y analiza ideas-patrones obtenidos desde los datos y valida correlaciones inusuales que son

entregadas a los expertos en la materia; un modelo ML clasifica todas las transacciones en normales o anormales; aunque es necesario tener transacciones revisadas manualmente que sirven como muestra para recorrer y recolectar las transacciones valiosas en forma más eficiente y rápida (Alvarado Salazar, 2022; Labanca et al., 2022; Sanchez-Romero & Llerena-Izquierdo, 2023; Sánchez Guzmán, 2021).

Se posible conocer detalles específicos acerca del uso de ML contra el lavado de dinero en organizaciones a nivel de artículos científicos, esta investigación se enfoca en trabajos sobre uso ML en transacciones o empresas que obtienen fondos ilegales; no se aborda la fuente de obtención del dinero, no se aborda el engaño a las personas, aunque se conoce que las transacciones se generan varias veces y en valores pequeños para que los criminales accedan a los fondos en forma segura.

Se propone generar un modelo en Machine Learning para el análisis de prevención de lavado de dinero en Ecuador, para esto se adopta componentes y algoritmo desde los artículos científicos a nivel de Ecuador y del mundo.

2. REVISIÓN DE LITERATURA

2.1. Machine Learning (ML)

Es una sub-rama de la Inteligencia Artificial (IA), que utiliza estadísticas, técnicas lógicas y matemáticas para asistir a una máquina en aprender e inicia con datos sin programación y con una primicia general de conclusión que usa muestras de datos; en otras palabras, una computadora inicia el aprendizaje con su propia experiencia mediante la técnica de IA en el reconocimiento de patrones (Alvarado-Salazar & Llerena-Izquierdo, 2022)(Puga Paredes, 2023); las conclusiones se basan en el entrenamiento previo sobre los datos. Existen dos tipos de algoritmos básicos de ML que son supervisados y no supervisados, aunque hay autores que presentan más tipos ML (Das et al., 2022).

2.2. Algoritmos de Machine Learning

Algoritmos supervisados: son algoritmos que utilizan solo datos etiquetados para entrenar y alcanzar una clasificación más eficiente; algunos algoritmos son Support vector machine (SVM), k-nearest neighbor (k-NN), Random Forest, Naive Bayesian Network, Decision tree, Multi-Layer Perceptron (MLP), Artificial Neural Network (ANN).

Algoritmos no supervisados: son algoritmos que entrenan con datos que están sin etiquetar ni clasificar, funciona en la extracción de características sobre los datos, no conoce ninguna etiqueta-clase sobre los datos; algunos algoritmos son K-means, K- medoid, Rough C-Means (RCM), Fuzzy C-Means (FCM), Rough-Fuzzy C-Means (RFCM) (Das et al., 2022).

2.3. Lavado de dinero (LD)

Esta actividad está relacionada a otros delitos como suministro de medicamentos o fraude; detectar en forma eficiente los movimientos de esta actividad fortalece la prevención y el juicio de esta clase de delito; además hay varias categorías de lavado de dinero como juegos de azar, uso de propiedades o creación de negocios para encubrir la verdadera fuente del dinero (Zand et al., 2020).

El LD inicia como un delito determinante o subyacente, y finaliza con fondos disponibles en forma segura o alarma mínima; el LD es ese proceso de convertir el dinero delictivo en activos; hoy el blanqueo de capitales es un verdadero problema económico con daños a la sociedad e

instituciones financieras porque los orígenes de los fondos se ocultan y parecen fondos legales (Oztas et al., 2022).

2.4. Casos de prevención del lavado de dinero mediante ML

El sistema de aprendizaje detecta transacciones anómalas desconocidas y entrega patrones no vistos anteriormente mediante algoritmo Random Forest, utiliza etiquetas clasificadas por especialistas en lavado de dinero para optimizar la tasa de detección; el sistema pre-procesa las transacciones, además la agregación se realiza porque que los patrones en lavado de dinero están en varias transacciones en un período de tiempo; el modelo da una puntuación de anomalía a las transacciones, y la selección para enviar al analista financiero (Labanca et al., 2022).

Las transacciones son cifradas por las empresas y los auditores clasifican las transacciones con un sistema informático en algoritmo LSTM para detectar el blanqueo de capitales, además cumplen el esquema de seguridad y evitar que terceros infieran sobre las transacciones, para la seguridad utilizan el cifrado de secreto compartido (Zand et al., 2020).

El enfoque revisa los algoritmos ML que utilizan los bancos de Dinamarca en la lucha contra el lavado de dinero como KNN y Random Forest, además entregan una revisión técnica de los estadísticos e ingenieros (Jensen & Iosifidis, 2023).

Los autores, revisaron trabajos que detectan el lavado de dinero y descubrieron que utilizan SVM y Random Forest, además obtuvieron los enfoques como: riesgos, anomalías, análisis de comportamiento, casos históricos, análisis de gráficos (Oztas et al., 2022).

El modelo gestiona los falsos positivos, es decir, las transacciones que se encuentran bloqueadas por una falsa alarma; mantiene un menor tiempo de procesamiento; utiliza algoritmos de ML como NB, SVM y DT; aquí SVM es mejor que los otros algoritmos (Alkhalili et al., 2021).

Otros trabajos científicos describen el ML aplicado en detección de fraude y lavado de dinero, y expresan la necesidad de combinarlos, los algoritmos supervisados mantienen un alto rendimiento al momento de detectar fraudes conocidos y que es necesario las técnicas no supervisadas para descubrir nuevos patrones de lavado de dinero (Chen et al., 2018).

El modelo extrae perfiles locales, globales y/o temporales, la captura es entregada a los analistas que revisan los comportamientos, se destaca la seguridad en los sistemas de detección de fraude (Z. Chen, L. D. Van Khoa, 2020).

3. METODOLOGÍA

Para la extracción y análisis de artículos científicos, se exploran bibliotecas digitales y se adopta el proceso de (Sharma & Singh, 2018) que consiste de cinco etapas: Preguntas de investigación, Fuentes de datos, Inclusión-Exclusión, Extracción de datos y Síntesis de datos. A continuación, se describen las etapas. Figura 1.

Figura 1.

Proceso de revisión sistemática

Etapa 1. Preguntas de investigación (PI): Las PI tienen un papel importante para solventar la estrategia de búsqueda, la extracción y el análisis de datos.

Etapa 2. Fuente de datos y búsqueda: Las fuentes de datos son las bibliotecas ACM Digital Library y IEEE Xplore. La búsqueda tiene las palabras claves relevantes y relacionadas a las PI y Machine Learning y lavado de dinero: “Machine Learning money laundering”, “money laundering”.

Etapa 3. Inclusión y Exclusión: Aquí se establecen los criterios de inclusión y exclusión para obtener con precisión la calidad de artículos. Criterios de inclusión son: Los artículos disponibles en formato completo, artículos en idioma inglés, publicados desde el año 2020 al 2023. Criterios de exclusión son: Artículos por pagar, artículos resumen, capítulos de libros.

Etapa 4. Extracción de datos: Esta se basa en los contenidos de los artículos seleccionados y en una hoja electrónica se plasman las propiedades que corresponden a las preguntas de investigación.

Etapa 5. Síntesis de datos: Aquí se realiza la recopilación, creación de estadísticas y análisis de las respuestas a las preguntas de investigación. Se utiliza la deducción y conclusión.

Para elaborar un modelo teórico basado en tecnología Machine Learning y que ayude en la prevención de lavado de dinero en Ecuador, se adopta un algoritmo ML de acuerdo con la extracción y análisis de artículos científicos que recomienden el algoritmo más apropiado; se establece los pasos para ejecutar el modelo, se establece las características del conjunto de datos, se realiza una representación gráfica del modelo y un pequeño ejemplo del conjunto de datos.

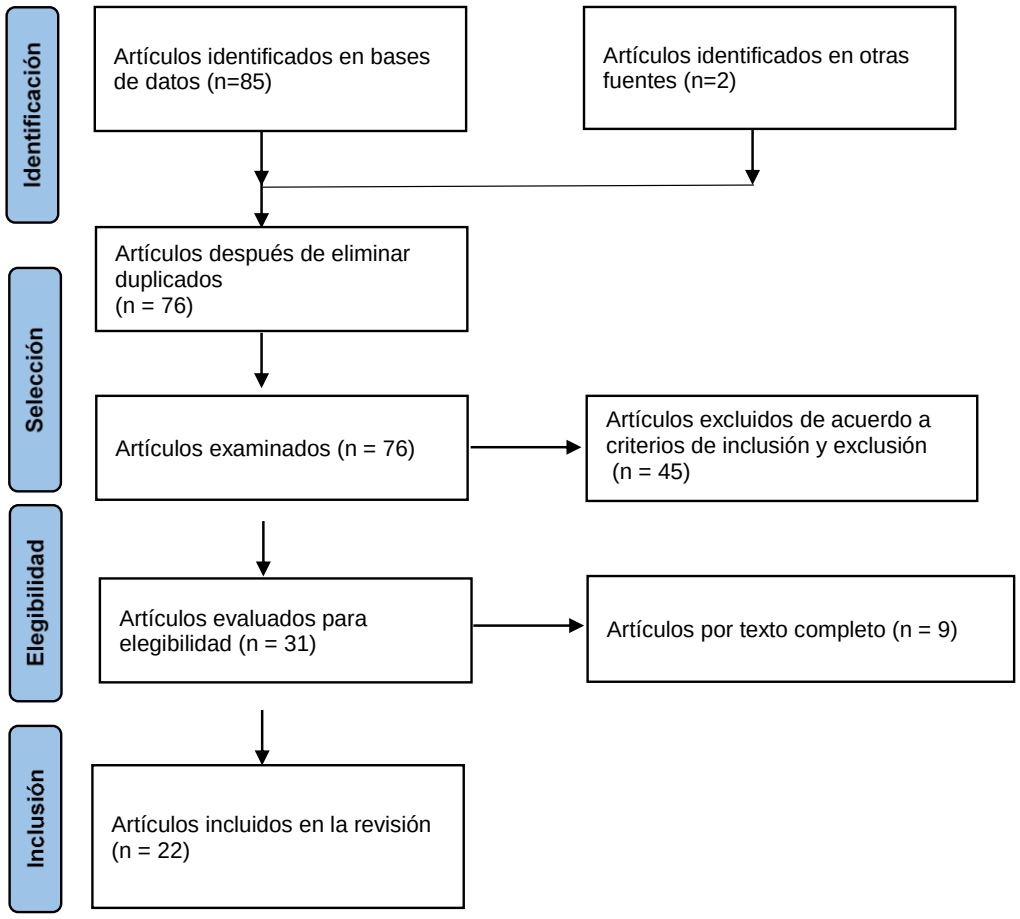
4. RESULTADOS

Extraer artículos científicos sobre Machine Learning y Lavado de dinero mediante la exploración en bibliotecas digitales.

Para extraer artículos científicos sobre Machine Learning y Lavado de dinero mediante la exploración en bibliotecas digitales, se utiliza el método PRISMA formado por cuatro fases (identificación, selección, elegibilidad e inclusión). La primera recolección de artículos es 85 documentos de IEEE y ACM, a esto se suman 2 de Google Scholar. Después de eliminar los artículos duplicados entre las fuentes, se seleccionan y se examinan 76 artículos. Después se aplican los criterios de inclusión/exclusión y esto hace que se eliminen 45 artículos para quedar elegibles los 31 documentos restantes. Se realiza la lectura completa de los 31 artículos, y se eliminan 9 artículos por no ser apegados a los objetivos de esta investigación. Finalmente son 22 artículos que sirven para el análisis en una hoja electrónica. Figura 2.

Figura 2.

PRISMA



La Tabla 1 proporciona la cantidad de artículos publicados sobre Machine Learning y Lavado de Dinero, durante los años 2020 al 2023. Los años especifican la distribución de los artículos e indica que los investigadores hacen esfuerzos continuos para mejorar las predicciones o análisis utilizando la tecnología ML. Los principales orígenes de publicación son IEEE Xplore y ACM Digital Library. Además, esto indica que los problemas de lavado de Dinero que utilizan perspectivas de ML están captando impulso. Los 22 artículos son la base para responder las preguntas de investigación.

Tabla 1. Referencias

No	Artículos seleccionados
2	(Zand et al., 2020), (Guevara et al., 2020),
7	(Alkhalili et al., 2021), (Mahootiha et al., 2021), (Tai, 2021), (Kute et al., 2021), (Chen & Soliman, 2021), (Raiter, 2021), (Domashova, 2021)
7	(Labanca et al., 2022), (Oztas et al., 2022), (Das et al., 2022), (Wang, 2022), (Mohammed & Thomas, 2022), (Li et al., 2022), (Eddin et al., 2022)
6	(Jensen & Iosifidis, 2023), (Adam, 2023), (Thommandru et al., 2023), (Cherkaoui & En-Naimi, 2023), (Met et al., 2023), (Lokanan, 2023)

Fuente: Autor.

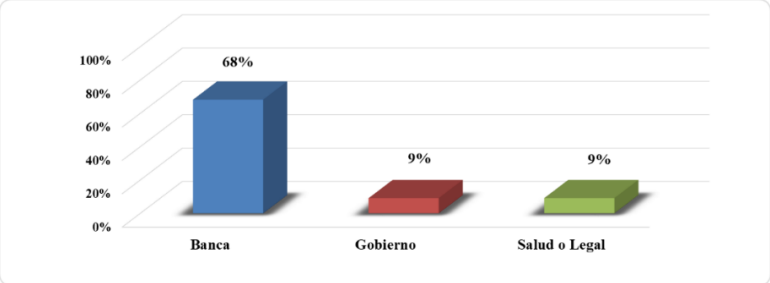
Analizar los resultados sobre la tecnología Machine Learning en la prevención del lavado de dinero mediante la deducción e inducción de los análisis de artículos científicos.

¿En qué sectores se utiliza ML contra lavado de dinero?

En los 22 artículos; el 68% (15 referencias) dirigen la investigación hacia la banca, otro 9% (2 referencias) dirigen la investigación hacia el gobierno, y otro 9% (2 referencias) dirigen la investigación hacia la salud o base legal. Aquí, 15 artículos seleccionados especificaron que los algoritmos ML se utilizan para el monitoreo de las transacciones financieras porque esto es una obligación para los bancos, además el monitoreo a través de ML es bien complementado de acuerdo con reglas para disminuir los falsos positivos. Figura 3.

Figura 3.

Sectores



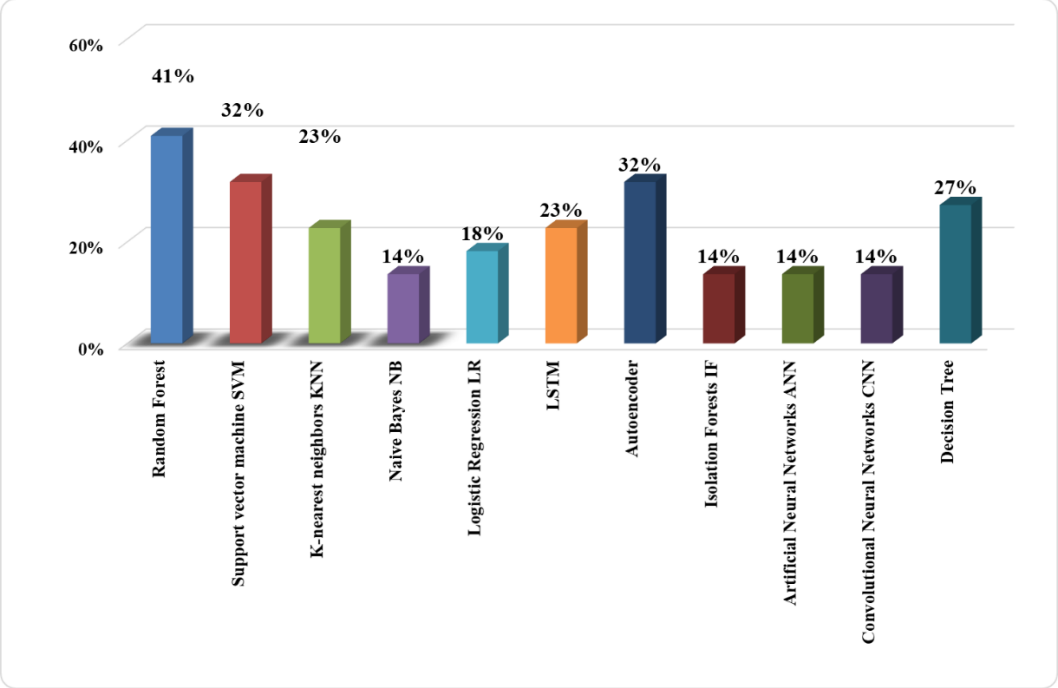
¿Qué algoritmos ML se utilizan contra el lavado de dinero?

En los 22 artículos; el 41% (9 referencias) utilizan Random Forest, el 32% (7 referencias) utilizan Support Vector Machine SVM, el 23% (5 referencias) utilizan K-Nearest Neighbors KNN, el 14% (3 referencias) utilizan Naive Bayes NB, el 18% (4 referencias) utilizan Logistic Regression LR, el 23% (5 referencias) utilizan LSTM, el 32% (7 referencias) utilizan Autoencoder, el 14% (3 referencias) utilizan Isolation Forests IF, el 14% (3 referencias) utilizan Artificial Neural Networks ANN, el 14% (3 referencias) utilizan Convolutional Neural Networks CNN, y el 27% (6 referencias) utilizan Decision Tree. Figura 4.

Aquí, el algoritmo Random Forest es el más utilizado, de acuerdo a (Eddin et al., 2022) este algoritmo proporciona un mejor rendimiento. De acuerdo a (Raiter, 2021) el algoritmo Random Forest tiene puntuación de 0.99 y mantiene una puntuación de precisión más alta.

Figura 4.

Algoritmos ML

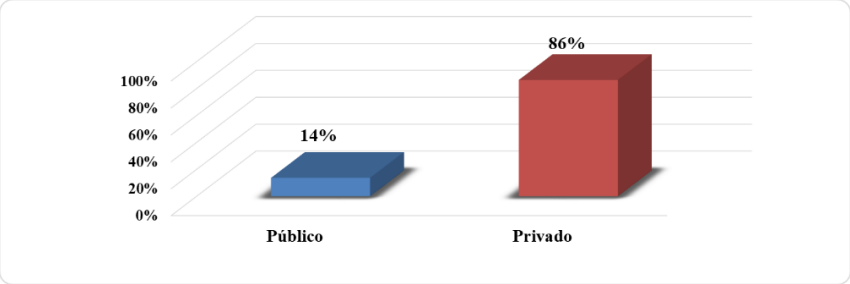


¿En qué tipos de entidades se utiliza ML contra el lavado de dinero?

En los 22 artículos; el 14% (3 referencias) se utilizan en entidades del sector público, y el 86% (19 referencias) se utilizan en entidades del sector privado. Figura 5.

Figura 5.

Tipos de entidades



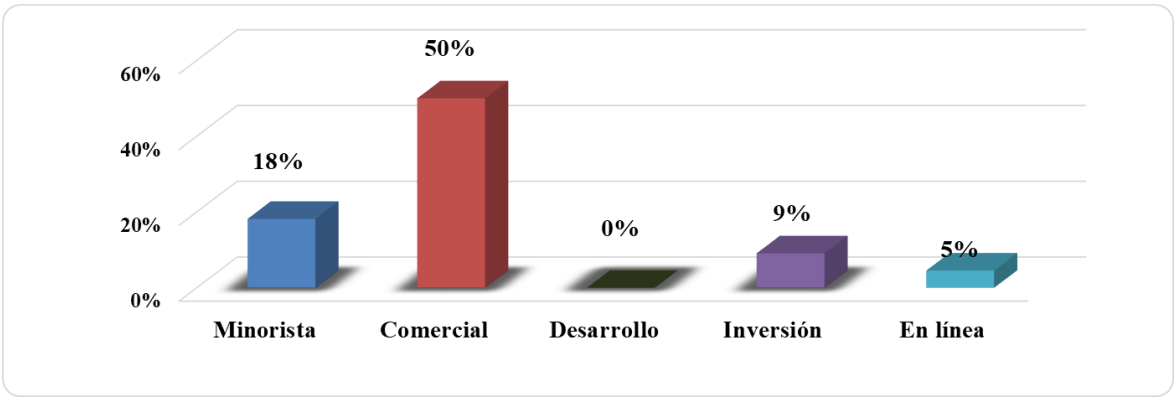
¿En qué clases de entidades utilizan ML contra el lavado de dinero?

En los 22 artículos; el 18% (4 referencias) se utilizan en entidades Minoristas, el 50% (11 referencias) se utilizan en entidades Comerciales, ninguna se utiliza en entidades de Desarrollo, el 9% (2 referencias) se utiliza en entidades de Inversión, el 5% (1 referencia) se utiliza en entidades En línea. Figura 6.

El lavado de dinero utiliza más el sector comercial para tener fondos disponibles que sean seguros y mínimo riesgo; esto también se llama blanqueo de capitales que genera un problema económico con daños a la sociedad. El origen de esos fondos son manipulados para aparentar legalidad, y utilizan tres métodos que son colocación, estratificación e integración (Oztas et al., 2022).

Figura 6.

Clases de entidades financieras



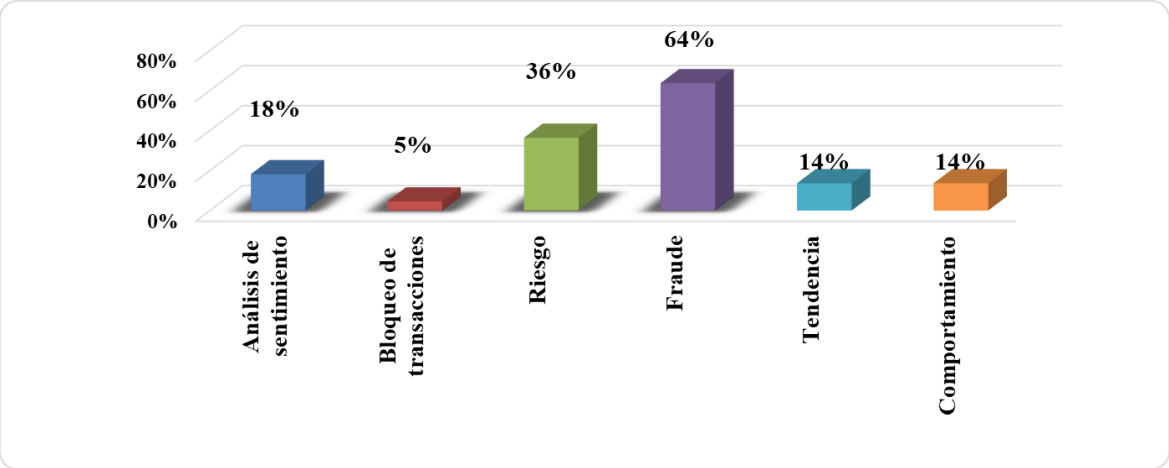
¿Qué actividades realizan los algoritmos ML contra el lavado de dinero?

En los 22 artículos; el 18% (4 referencias) se realizan el Análisis de sentimiento, el 5% (1 referencia) realiza el Bloqueo de transacciones, el 36% (8 referencias) analiza el Riesgo, el 64% (14 referencias) analiza el Fraude, el 14% (3 referencias) analiza las Tendencias, el 14% (3 referencias) analiza el Comportamiento. Figura 7.

La utilización de la tecnología ML ayuda a combatir, monitorear y analizar la actividad financiera ilícita; algunos expertos combinan redes neuronales con ML, obtienen buenos resultados para analizar conjuntos de datos muy grandes y optimizar el análisis de fraudes y riesgos.

Figura 7.

Clases de entidades financieras



¿Qué técnicas realizan los algoritmos ML contra el lavado de dinero?

En los 22 artículos; el 73% (16 referencias) se realizan la Clasificación, el 27% (6 referencias) realizan la Segregación, el 23% (5 referencias) realizan la Selección, el 50% (11 referencias) realizan la Predicción, el 14% (3 referencias) realizan el Filtrado, el 36% (8 referencias) realizan el Análisis, el 18% (4 referencias) realizan la Regresión. Figura 8.

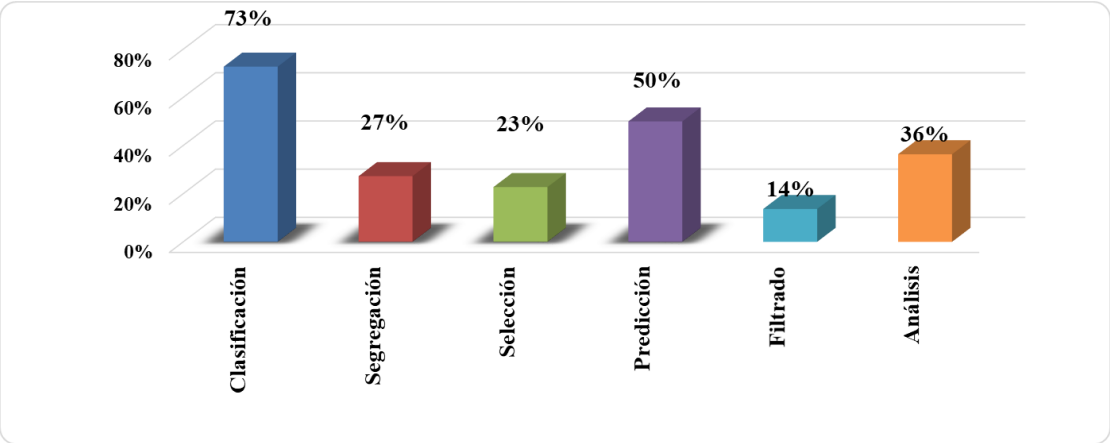
La Clasificación la ejecutan los algoritmos Random Forest, Support Vector Machine SVM, Naive Bayes NB, K-nearest neighbors KNN.

La Predicción la ejecutan los algoritmos Random Forest, LSTM, Support Vector Machine SVM, K-nearest neighbors KNN.

Para el Análisis se utilizan los algoritmos LSTM, Support Vector Machine SVM, Isolation Forests IF.

Figura 8.

Técnicas de ML



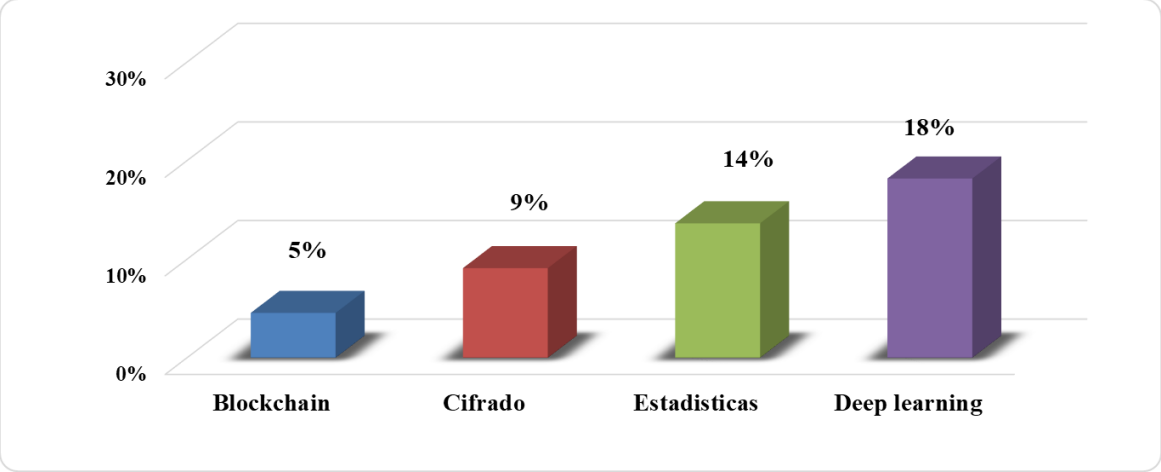
¿Se utilizan otras herramientas con los algoritmos ML contra el lavado de dinero?

En los 22 artículos; el 5% (1 referencia) también utiliza Blockchain, el 9% (2 referencias) también utiliza Cifrado, el 14% (3 referencias) también utiliza Estadísticas, y 18% (4 referencias) también utiliza Deep Learning. Figura 9.

Deep Learning se utiliza con otros algoritmos ML como: Support vector machine SVM, Logistic Regression LR, LSTM, Autoencoder, Artificial Neural Networks ANN, Convolutional Neural Networks CNN (Das et al., 2022). Decision Tree (Mohammed & Thomas, 2022), K-Nearest Neighbors KNN (Mahootiha et al., 2021), Logistic Regression LR (Kute et al., 2021).

Figura 9.

Técnicas de ML



De acuerdo al análisis de los resultados sobre la tecnología Machine Learning en la prevención del lavado en los 22 artículos seleccionados, **el algoritmo ML más apropiado es Random Forest (RF)** en 9 artículos, por las siguientes razones:

> Es una estructura para categorizar datos en diferentes clases; inicia en el nodo raíz, cada punto de datos traspasa diferentes ramas del árbol, de acuerdo con condiciones determinadas para cada nodo, hasta alcanzar a un nodo hoja.
>
> RF es un algoritmo que sigue la ruta del punto de datos por medio del árbol para establecer la condición que cumple o no para clasificar el punto atravesado.
>
> RF utiliza un conjunto de múltiples árboles de decisión que lo lleva a formar un bosque aleatorio, para obtener una predicción se ejecuta un promedio de la predicción de cada árbol individual. Durante el entrenamiento se obtienen nuevas etiquetas, RF se entrena de nuevo en consecuencia y se obtienen predicciones con los vectores sin etiquetar más datos.

Elaborar un modelo teórico en análisis de prevención de lavado de dinero en Ecuador basado en un algoritmo de ML

La detección sobre el blanqueo/lavado de capitales/dinero requiere datos etiquetados, una forma de mantener etiquetas es darle a cada transacción la revisión manual por especialistas, pero esto no es factible. Por este motivo, se utiliza ML, que consiste en analizar las actividades más sospechosas clasificadas mediante puntuación de transacciones extrañas; después, el modelo se entrena con los datos retroalimentados de los especialistas (un conjunto de datos etiquetados) para que sea posible escoger puntos de datos adicionales para su análisis.

Se selecciona el clasificador **RF** porque se basa en la puntuación de probabilidad, el primer paso es la selección de los vectores más extraños; el segundo paso es la selección de las transacciones agregadas menos extrañas. Aquí, se puede aprovechar la mayor precisión del clasificador **RF** para robustecer la información que está dentro del conjunto de datos etiquetado y etiqueta-califica en forma automática cada transacción del conjunto (Labanca et al., 2022).

La primera actividad en el flujo de trabajo del modelo es adicionar las transacciones sin procesar que pertenecen a un período de tiempo para generar características que representen “vectores de alto nivel” que obtengan el perfil de comportamiento de un cliente. El modelo puede entrenarse con estos “vectores de alto nivel” generados desde los datos históricos. Luego de la

fase de entrenamiento, el modelo puede calcular una puntuación de transacción extraña para cada nuevo vector mediante el modelo supervisado. Luego se utiliza una estrategia de elección específica mediante la puntuación de transacción extraña para elegir los vectores que se remiten al especialista para su revisión. La cantidad de transacciones enviadas por cada día para la revisión es un parámetro del modelo; de acuerdo a los recursos que un banco pueda destinar a este proceso. El especialista analiza las transacciones para establecer si son extrañas o no. Después, estas transacciones se guardan como etiquetas en el conjunto de datos. Las etiquetas inspeccionadas contribuyen a un conjunto auténtico de datos etiquetados, este conjunto es la entrada para el componente del modelo. Luego, el componente del modelo selecciona de manera permanente los datos que revisará el especialista.

De acuerdo a (Labanca et al., 2022), los conjuntos de datos financieros son desequilibrados y contienen un promedio de 1% transacciones extrañas. Existen varios patrones sospechosos sugeridos por el GAFI (GAFI, 2023), que es un organismo intergubernamental que regula combate el lavado de dinero. Las transacciones extrañas tienen la misma tendencia que fortalecen el concepto que toda transacción extraña está oculta dentro del conjunto de datos.

Algunos de los tipos de transacciones extrañas son: a) Transacciones pequeñas pero habituales concebidas en un pequeño período de tiempo. b) Transacciones con montos sin decimales de compras o ventas en una cuenta. c) Valores por compra o venta en un momento inusual. d) Retiro de grandes montos sin fundamento comercial. e) Monto grande en transferencia en períodos cortos.

Las transacciones financieras son de carácter privado, por esta razón existen pocas estadísticas publicadas sobre los servicios financieros, en especial sobre las transferencias de dinero móvil. Algunos investigadores proponen conjuntos de datos sintéticos desarrollados por medio de algún simulador, que son datos agregados basados en datos privados que reproducen la actividad normal de las transacciones y tienen comportamientos dañinos para verificar la efectividad del algoritmo que detecte el fraude (Raiter, 2021).

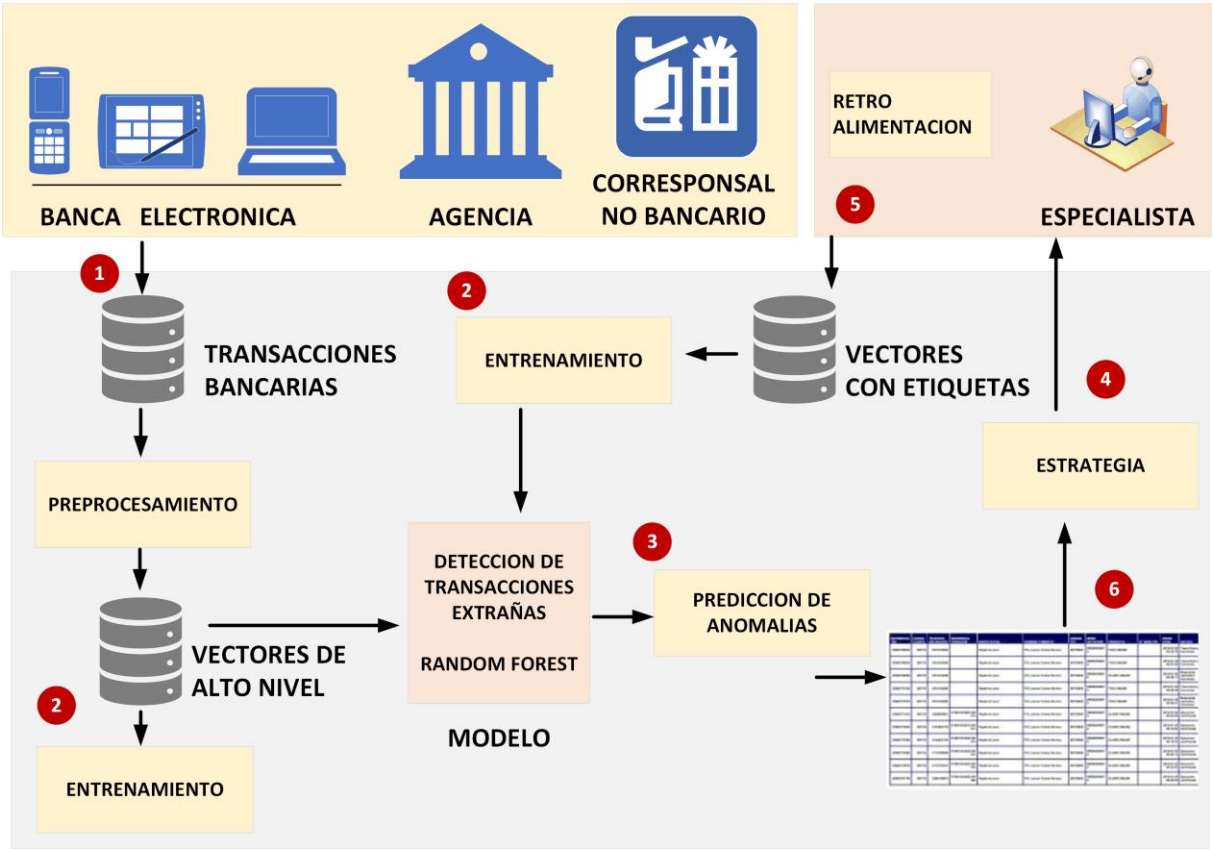
Esta investigación propone algunas de las características que el conjunto de datos debería tener y son: El valor de la transacción en moneda nacional al país. El nombre del cliente dueño de la transacción. El saldo original antes de la transacción. El nuevo saldo después de la transacción. El nombre del cliente beneficiario. El banco receptor del saldo original antes del movimiento. Nuevo saldo del beneficiario después de la transacción. Si acaso la transacción tiene actividad

fraudulenta en el conjunto de datos, que intenta tener control de las cuentas o intenta transferir dinero a otra cuenta y luego hace retiros. Si acaso es una transferencia de gran valor desde una cuenta hacia otra.

En la figura 10, los pasos principales están en círculos rojos, el primero es la agregación de transacciones históricas en vectores de alto nivel; el segundo en dos partes, es el entrenamiento del modelo de aprendizaje con los vectores de alto nivel, y además con la retroalimentación recolectada en el cuarto paso; el tercer paso es la puntuación de transacciones extrañas de los vectores de alto nivel mediante el modelo; el cuarto paso es determinación de estrategias de selección; el quinto paso es la obtención de las etiquetas del especialista; el sexto paso es la agregación de las puntuaciones, es decir, la predicción de transacciones extrañas o anómalas.

Figura 10.

Modelo de prevención en lavado de dinero.



La figura 11, se presenta del lado izquierdo (color verde) la agregación de las transacciones sin procesar que pasan a ser vectores de alto nivel en el lado derecho (color amarillo). Se muestra la secuencia de transacciones sin procesar, y los vectores agregados de alto nivel.

El vector contiene y presenta el nombre del cliente, fecha del conjunto de transacciones, la cantidad de transacciones, el promedio de valores y la suma de los valores de las transacciones. En este ejemplo se presentan 12 transacciones históricas, luego de la aplicación del modelo estas transacciones resultan en tres vectores de alto nivel.

Figura 11.

Ejemplo de datos y vectores

TRANSACCIONES

TRANSACCION	CLIENTE	FECHA	MONTO	TIPO
0001	AAAA	2/1/2024	428.00	AHORRO
0002	AAAA	2/1/2024	722.00	AHORRO
0003	AAAA	2/1/2024	116.00	CORRIENTE
0004	AAAA	2/1/2024	783.00	AHORRO
0005	AAAA	2/1/2024	215.00	CORRIENTE
0006	BBBB	4/1/2024	87.00	AHORRO
0007	BBBB	4/1/2024	938.00	CORRIENTE
0008	BBBB	4/1/2024	683.00	AHORRO
0009	BBBB	4/1/2024	690.00	AHORRO
0010	CCCC	10/1/2024	689.00	CORRIENTE
0011	CCCC	10/1/2024	78.00	AHORRO
0012	CCCC	10/1/2024	844.00	CORRIENTE

VECTORES DE ALTO NIVEL

CLIENTE	AAAA
FECHA	2/1/2024
CONTADOR	5
PROMEDIO	452.8
SUMA	2264

CLIENTE	BBBB
FECHA	4/1/2024
CONTADOR	4
PROMEDIO	599.50
SUMA	2398.00

CLIENTE	CCCC
FECHA	10/1/2024
CONTADOR	3
PROMEDIO	537.00
SUMA	1611.00

El criterio de clasificación en el árbol es producto de una partición recursiva en el conjunto de datos de entrenamiento original; aquí el nodo principal que es el vector de alto nivel se divide en nodos subsiguientes (transacciones) que son excluyentes de un número finito de transacciones, y estos dependen de los valores encontrados en las transacciones recopiladas.

En una variable X y un valor c, se define una división remitiendo todas las transacciones con valores de X menores o iguales que c al lado izquierdo y todas las transacciones restantes al lado derecho. El algoritmo utiliza el promedio de 2 valores adyacentes para calcular c. Entonces, una variable con N valores diferentes proporciona hasta N–1 divisiones potenciales del nodo principal.

5. DISCUSIÓN

Finalmente son 22 artículos que sirven para el análisis, y en contra el lavado de dinero se obtuvo lo siguiente: El mayor sector que utiliza ML es la Banca, el primer algoritmo ML que utilizan es RF, las mayores entidades que utilizan ML son Comerciales, la primera actividad que se realiza es el Fraude, la primera técnica que se realiza es la Clasificación, la mayor herramienta utilizada con algoritmos ML es Deep Learning.

El modelo que se propone se basa en un algoritmo supervisado que es Random Forest, de acuerdo a la literatura este optimiza la capacidad de cualquier modelo o sistema para obtener predicciones futuras; los modelos supervisados pueden generar predicciones más precisas que los modelos no supervisados. El algoritmo Random Forest es una forma de caja negra y funciona como un árbol a través de reglas cuya predicción se entiende con el camino que conduce a la clasificación. En esta investigación se adopta Random Forest porque de acuerdo a los artículos científicos (Labanca et al., 2022), (Jensen & Iosifidis, 2023), (Oztas et al., 2022), (Das et al., 2022), (Li et al., 2022), (Cherkaoui & En-Naimi, 2023), (Lokanan, 2023), (Raiter, 2021), (Eddin et al., 2022), RF obtiene el mejor rendimiento en relación con otros algoritmos. El segundo algoritmo más utilizado contra lavado de dinero es SVM, de acuerdo a los artículos científicos (Oztas et al., 2022), (Das et al., 2022), (Alkhalili et al., 2021), (Wang, 2022), (Tai, 2021), (Guevara et al., 2020), (Raiter, 2021), los modelos se mejoran utilizando el algoritmo SVM, y obtiene mayor precisión por las funciones y métodos que posee este algoritmo.

En este artículo sólo propone el diseño del modelo; se excluyeron tiempos de desarrollo o implementación para el modelo, se excluyeron los recursos humanos que se necesitarían para el desarrollo del modelo, se excluye los costos financieros que se necesitarían para la implementación, también se excluye algún prototipo. Algunas limitaciones en el trabajo de este artículo, es la lectura de artículos en menor cantidad, pero lo relevante es que son artículos científicos de ACM Digital Library y IEEE Xplore; otra limitante es el conjunto de transacciones bancarias porque no son datos de libre disponibilidad por las leyes de sigilo bancario en Ecuador. Los artículos científicos revisados crearon sus propios conjuntos de datos para el entrenamiento y prueba en los prototipos que presentaron. Utilizaron conjuntos de datos que son datos falseados de datos del mundo real, si acaso son de tiempo muy limitado y pocos días de transacciones. En cambio, un conjunto de datos más grande permitirá estudiar-analizar la maniobra del sistema a través del tiempo.

6. CONCLUSIÓN

El aumento de comercio, fabricación y consumo de productos ilegales tiene un flujo de dinero, los algoritmos Machine Learning pueden examinar las transacciones en tiempo real que entran al banco, y verificar si la transacción es un fraude o es auténtica. Las transacciones pueden ser marcadas como ilegales, y remitirse para investigaciones en futuro inmediato. Las operaciones que son sospechosas se registran por medio de un monitoreo continuo y detección/intercepción de estas transacciones.

Los modelos de ML optimizan la detección del lavado de dinero en las transacciones de rutina, los resultados de esta investigación sugieren que **Random Forest** es mejor en entornos de detección de lucha contra el lavado de dinero en el mundo real. Por la precisión e interpretación, es muy útil en la arquitectura contra el lavado de dinero en empresas financieras. Le siguen SVM y Árbol de Decisión.

El lavado de dinero mantiene características complicadas, como la estratificación por parte de los blanqueadores que intercambian dinero entre diferentes bancos y cuentas de ahorros/corriente para ocultar sus fuentes. Para capturar ese patrón se necesita trabajo adicional, como el análisis en redes sociales, junto con el modelo propuesto en este artículo. Por supuesto, que existe mucho margen de mejora en el modelo que se propone. Además, se necesita un conjunto de datos muy grande con campos o atributos profundos, o un buen grado de granularidad en los datos.

REFERENCIAS

Adam, T. (2023). Anomaly Detection on Distributed Ledger Using Unsupervised Machine Learning. *2023 IEEE International Conference on Omni-Layer Intelligent Systems (COINS)*, *2020*, 1–4. https://doi.org/10.1109/COINS57856.2023.10189278

Alkhalili, M., Qutqut, M. H., & Almasalha, F. (2021). Investigation of Applying Machine Learning for Watch-List Filtering in Anti-Money Laundering. *IEEE Access*, *9*, 18481–18496. https://doi.org/10.1109/ACCESS.2021.3052313

Alvarado-Salazar, R., & Llerena-Izquierdo, J. (2022). Revisión de la literatura sobre el uso de Inteligencia Artificial enfocada a la atención de la discapacidad visual. *Revista InGenio*, *5*(1), 10–21. https://doi.org/https://doi.org/10.18779/ingenio.v5i1.472

Alvarado Salazar, R. E. (2022). *Inteligencia artificial con enfoque a la discapacidad visual: un mapeo sistemático*.

Asamblea-Ecuador. (2023). *Ley Prevención de Lavado de Activos y del Financiamiento De Delitos*. 1–12.

Chen, Z., & Soliman, W. M. (2021). Variational Autoencoders and Wasserstein Generative Adversarial Networks for Improving the Anti-Money Laundering Process. *IEEE Access*, *9*, 83762–83785. https://doi.org/10.1109/ACCESS.2021.3086359

Chen, Z., Van 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: a review. *Knowledge and Information Systems*, *57*(2), 245–285. https://doi.org/10.1007/s10115-017-1144-z

Cherkaoui, R., & En-Naimi, E. M. (2023). A comparison of machine learning algorithms for credit card fraud detection. *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, 1–6. https://doi.org/10.1145/3607720.3607759

Das, P. K., A, D. V, Meher, S., Panda, R., & Abraham, A. (2022). A Systematic Review on Recent Advancements in Deep and Machine Learning Based Detection and Classification of Acute Lymphoblastic Leukemia. *IEEE Access*, *10*(July), 81741–81763. https://doi.org/10.1109/ACCESS.2022.3196037

Domashova, J. (2021). Usage learning for early detection Architectures for Artificial Intelligence : Eleventh Annual. *Procedia Computer Science*, *190*(2020), 184–192. https://doi.org/10.1016/j.procs.2021.06.033

Eddin, A. N., Bono, J., Apar, D., & Polido, D. (2022). *Anti-Money Laundering Alert Optimization Using Machine Learning with Graphs*. https://doi.org/https://doi.org/10.48550/arXiv.2112.07508

GAFI. (2023). *Grupo de Acción Financiera Internacional*. https://www.fatf-gafi.org/

Guevara, J., Garcia-Bedoya, O., & Granados, O. (2020). Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents. *Communications in Computer and Information Science*, *1277*, 72–88. https://doi.org/10.1007/978-3-030-61702-8_6

Jensen, R. I. T., & Iosifidis, A. (2023). Fighting Money Laundering With Statistics and Machine Learning. *IEEE Access*, *11*(January), 8889–8903. https://doi.org/10.1109/ACCESS.2023.3239549

Kute, D. V., Pradhan, B., Shukla, N., & Alamri, A. (2021). Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering – A Critical Review. *IEEE Access*, *9*, 82300–82317. https://doi.org/10.1109/ACCESS.2021.3086230

Labanca, D., Primerano, L., Markland-Montgomery, M., Polino, M., Carminati, M., & Zanero, S. (2022). Amaretto: An Active Learning Framework for Money Laundering Detection. *IEEE Access*, *10*, 41720–41739. https://doi.org/10.1109/ACCESS.2022.3167699

Li, Z., Zhang, Y., Wang, Q., & Chen, S. (2022). *Transactional Network Analysis and Money Laundering Behavior Identification of Central Bank Digital Currency of China*. *3*(3).

Lokanan, M. E. (2023). Predicting Money Laundering Using Machine Learning and Artificial Neural Networks Algorithms in Banks. *Journal of Applied Security Research*, *19*(1), 20–44. https://doi.org/10.1080/19361610.2022.2114744

Mahootiha, M., Golpayegani, A. H., & Sadeghian, B. (2021). Designing a New Method for Detecting

Money Laundering based on Social Network Analysis. *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, 1–7. https://doi.org/10.1109/CSICC52343.2021.9420621

Met, I., Erkoc, A., & Seker, S. E. (2023). Performance, Efficiency, and Target Setting for Bank Branches: Time Series With Automated Machine Learning. *IEEE Access*, *11*(January), 1000–1010. https://doi.org/10.1109/ACCESS.2022.3233529

Mohammed, H. N., & Thomas, S. (2022). Machine Learning Approach to Anti-Money Laundering : A Review. *2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development (NIGERCON)*, 1–5. https://doi.org/10.1109/NIGERCON54645.2022.9803072

Oztas, B., Cetinkaya, D., Adedoyin, F., & Budka, M. (2022). Enhancing Transaction Monitoring Controls to Detect Money Laundering Using Machine Learning. *2022 IEEE International Conference on E-Business Engineering (ICEBE)*, 26–28. https://doi.org/10.1109/ICEBE55470.2022.00014

Pérez González, R. F. (2021). *Softwares de penetración utilizados por los piratas informáticos: Una revisión sistemática (2015-2020)*.

Puga Paredes, J. L. (2023). *Mapeo sistemático sobre el seguimiento del aprendizaje de estudiantes mediante el uso de minería de datos educativos*.

Raiter, O. (2021). Applying Supervised Machine Learning Algorithms for Fraud Detection in Anti-Money Laundering. *Journal of Modern Issues in Business Research*, *1*(1), 14–26.

Rosero Tejada, L. F. (2021). *El Phishing como riesgo informático, técnicas y prevención en los canales electrónicos: Un mapeo sistemático*.

Sanchez-Romero, J., & Llerena-Izquierdo, J. (2023). Revisión de la literatura sobre el uso del aprendizaje profundo enfocado en sistemas de inspección ópticos automatizados para la detección de defectos superficiales en el sector de la manufactura. *Revista InGenio*, *6*(2), 1–19. https://doi.org/10.18779/ingenio.v6i2.680

Sánchez Guzmán, C. O. (2021). *Modelo de red segura en un entorno distribuido para la transferencia de datos con mecanismos básicos de seguridad*.

Sharma, P., & Singh, J. (2018). Systematic Literature Review on Software Effort Estimation Using Machine Learning Approaches. *2018 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, 43–47. https://doi.org/10.1109/ICNGCIS.2017.33

Tacuri López, I. L. (2021). *Acoso por medio de las tecnologías en las redes sociales durante tiempos de pandemia en Ecuador, una revisión sistemática*.

Tai, C. (2021). Identifying Money Laundering Accounts. *2019 International Conference on System Science and Engineering (ICSSE)*, 379–382.

Thommandru, A., Mone, V., Mitharwal, S., & Tilwani, R. (2023). Exploring the Intersection of Machine Learning , Money Laundering , Data Privacy , and Law. *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, 149–155. https://doi.org/10.1109/ICIDCA56705.2023.10099859

Toala Indio, Y. I. (2021). *Delitos informáticos frecuentes en el Ecuador: casos de estudio*.

Vera Navas, N. A. (2021). *Modelo de seguridad informática para riesgos de robo de información por el uso de las redes sociales*. http://dspace.ups.edu.ec/handle/123456789/20949

Wang, Z. (2022). Detection Mechanism of Money Laundering based on Random Walk and Skip-Grim Model. *2022 IEEE 5th International Conference on Electronic Information and Communication Technology (ICEICT)*, 444–448. https://doi.org/10.1109/ICEICT55736.2022.9909113

Yahaya, S. Z., Mohd Zailani, M. N., Che Soh, Z. H., & Ahmad, K. A. (2020). IoT Based System for Monitoring and Control of Gas Leaking. *Proceeding - 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, ICITAMEE 2020*, 122–127. https://doi.org/10.1109/ICITAMEE50454.2020.9398384

Yaya, M. H. B. M., Patchmuthu, R. K., & Wan, A. T. (2021). LPG Gas Usage and Leakage Detection Using IoT in Brunei. *2021 International Conference on Green Energy, Computing and Sustainable Technology, GECOST 2021*, 1–5. https://doi.org/10.1109/GECOST52368.2021.9538647

Z. Chen, L. D. Van Khoa, E. (2020). Artificial Intelligence in Financial Services. *Inf. Syst.*, *57*.

Zand, A., Orwell, J., & Pfluegel, E. (2020). A Secure Framework for Anti-Money-Laundering using Machine Learning and Secret Sharing. *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1–7. https://doi.org/10.1109/CyberSecurity49315.2020.9138889

Zia, U. U. R., Zulfiqar, M., Azram, U., Haris, M., Khan, M. A., & Zahoor, M. O. (2019). Use of Macro/Micro Models and Business Intelligence tools for Energy Assessment and Scenario based Modeling. *2019 4th International Conference on Emerging Trends in Engineering, Sciences and Technology, ICEEST 2019*. https://doi.org/10.1109/ICEEST48626.2019.8981691