

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERÍA Y ARQUITECTURA
Escuela Profesional de Ingeniería de Sistemas



Modelos de aprendizaje automático aplicados a la detección de transacciones sospechosas de lavado de activos en entidades financieras: Una revisión sistemática de la literatura

Por:

Amador Junior Galeano Villar
Zannier Noe Vargas Cisneros

Asesor:

Mg. Keyla De la Cruz Gonzales
Dr. Guillermo Mamani Apaza

Lima, diciembre 2019

**DECLARACIÓN JURADA
DE AUTORÍA DEL TRABAJO DE
INVESTIGACIÓN**

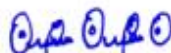
Mg. Keyla Dervith De La Cruz Gutierrez, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que el presente trabajo de investigación titulado: "MODELOS DE APRENDIZAJE AUTOMÁTICO APLICADOS A LA DETECCIÓN DE TRANSACCIONES SOSPECHOSAS DE LAVADO DE ACTIVOS EN ENTIDADES FINANCIERAS: UNA REVISIÓN SISTEMÁTICA DE LA LITERATURA" constituye la memoria que presentan los estudiantes Amador Junior Galeano Villar y Zannier Noe Vargas Cisneros para aspirar al grado de bachiller en Ingeniería de Sistemas, cuyo trabajo de investigación ha sido realizado en la Universidad Peruana Unión bajo mi dirección.

Las opiniones y declaraciones en este trabajo de investigación son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente declaración en la ciudad de Lima, al 03 de diciembre del año 2019.



Mg. Keyla Dervith de la Cruz Gutierrez


Modelos de aprendizaje automático aplicados a la detección de transacciones sospechosas de lavado de activos en entidades financieras: Una revisión sistemática de la literatura

Trabajo de investigación

Presentado para optar al grado de bachiller en Ingeniería de Sistemas

JURADO CALIFICADOR


Dra. Erika Inés Acuña Salinas
Presidente


Mg. Omar Leonel Loaiza Jara
Secretario


Ing. Fredy Aber Huanca Torres
Vocal


Mg. Herminio Paucar Curasma
Vocal


Mg. Keyla Dervith De la Cruz Gutierrez
Asesor

Lima, 02 de diciembre del 2019

Modelos de aprendizaje automático aplicados a la detección de transacciones sospechosas de lavado de activos en entidades financieras: Una revisión sistemática de la literatura

Amado Junior Galeano Villar¹, Zannier Noe Vargas Cisneros²

^{1,2} Universidad Peruana Unión, Lima, Perú
amadorgaleano@upeu.edu.pe
zanniervargas@upeu.edu.pe

Resumen. El lavado de activos es uno de los delitos que viene afectando a la economía del país. Grandes cantidades de dinero se lavan todos los años. Según Daniel Linares, intendente de Análisis Operativo de la UIF estimó que entre junio de 2016 y mayo de 2017 el monto investigado aumentó en 125%. Por esta razón, este estudio tiene como objetivo identificar modelos de aprendizaje automático propuestos, diseñados o implementados para el apoyo en la detección de transacciones sospechosas de lavado de activos en entidades financieras. Para lograr identificar los modelos de aprendizaje automático se realizó una revisión sistemática de la literatura de las investigaciones publicadas en las diferentes librerías digitales indexadas. De un total de 485 artículos revisados, se identificaron 20 artículos que hacen referencia a los modelos de aprendizaje automático. Cabe destacar que los modelos de aprendizaje automático son comúnmente utilizados para apoyar en la detección de transacciones sospechosas de lavado de activos por su adecuación al entorno cambiante, siendo esto una de sus ventajas sobre los sistemas tradicionales de monitorización. Actualmente existen diversidad de métodos, algoritmos y técnicas de aprendizaje automático aplicados para lograr este fin, siendo los algoritmos de agrupación los que mayormente se utilizan según los estudios seleccionados.

Palabras claves: Modelos de aprendizaje automático, algoritmos de aprendizaje, transacciones sospechosas, lavado de activos.

1 Introducción

Debido a las fuertes interconexiones entre la demanda, la inversión, el comercio y la productividad el incremento de la economía mundial tiene un impacto positivo, ya que el desarrollo de las diferentes actividades económicas que cada empresa desarrolla presentan nuevas formas de operar dentro del sistema financiero, producto de la expansión de sus negocios a otros países con el fin de intercambiar bienes y servicios para que ambos puedan enriquecerse aprovechando sus ventajas. El intercambio comercial y monetario se da dentro del desarrollo de sus actividades normales y en la mayoría de los casos el pago se realiza mediante una transacción en entidades financieras puesto que estas pueden estar presentes en los diferentes países.

Dentro de las transacciones monetarias, uno de los delitos financieros como el lavado de activos se hace presente y ha evolucionado con el pasar de los años y cada día las organizaciones criminales buscan mejorar e innovar sus estrategias para cometer actos ilícitos. Es así que en algunas entidades financieras existe una predisposición de evadir la ley y colaborar con este delito. Además, las transacciones financieras y el control de las mismas están en manos de algunas empresas multinacionales que dictan sus propias reglas, se mantienen al margen o contra las normas y leyes que regulan sus operaciones, tratando de evadir procedimientos y controles del sistema financiero. Por esta razón se crea la Superintendencia de Banca Seguros y AFP(SBS) para regular las actividades de las diferentes entidades financieras, y por otro lado a través de la Unidad de Inteligencia Financiera(UIF) juntamente con el Banco Interamericano de Desarrollo(BID) presentar informes de la Evaluación Nacional de Riesgos de Lavados de Activos y Financiamiento del Terrorismo, para conocer los riesgos y vulnerabilidades que perjudican al sistema financiero mediante la evaluación de las características de las entidades financieras peruanas, a fin de detectar las debilidades y métodos que faciliten la ejecución de operaciones de lavado de activos o financiamiento del terrorismo en el Perú. Es así que las transacciones sospechosas de lavado de activos recibidos por la Unidad de Inteligencia Financiera en el periodo comprendido entre enero de 2010 y julio de 2019 ha recibido 70,332 Reportes de Operaciones Sospechosas (ROS) por parte de los sujetos obligados a reportar.

Por otro parte, la SBS viene trabajando un proyecto de ley que establece un proceso de licenciamiento para las casas de cambio y casas de préstamos. Actualmente estos negocios solo se abren con una licencia municipal y nadie investiga los antecedentes de los dueños ni el origen del dinero.

Es por eso que la implementación de nuevas tecnologías y métodos de la ciencia de datos toma un rol muy importante en los últimos años para combatir este delito porque al igual que la aplicación de estos en el mundo empresarial ha brindado grandes contribuciones y aportes para la toma de decisiones con respecto a sus objetivos estratégicos, se tiene claro que también puede ser utilizado para apoyar en la mitigación del lavado de activos abarcando un mayor campo de acción y haciendo que estas soluciones sean adaptativas a las distintas formas de operar de los criminales. Dentro de la ciencia de datos, el campo del machine learning nos brinda modelos que se pueden usar con el fin de aumentar la capacidad de detección de las transacciones sospechosas, por ejemplo, encontramos uno que fue planteado por Zengan [3], el

modelo CBLOF el cual combina la agrupación no supervisada basada en la distancia y la detección local de valores atípicos. En este modelo la agrupación se realiza con el fin de preprocesar los datos para la identificación de la anomalía. En lo que respecta a la naturaleza del lavado de activos, el modelo de agrupación que se seleccione debe ser capaz de generar el número de grupos automáticamente y todos los grupos deben clasificarse de acuerdo con el número de los componentes en cada uno. Por otro lado, Camino *et al.* [6] usan un modelo que combina tres algoritmos que son Bosque de Aislamiento, OCSVM y GMM, todo esto con el fin de una mejor precisión en la detección de las transacciones sospechosas. En este modelo el algoritmo Bosque de Aislamiento ayuda a extraer reglas que deciden cuándo una cuenta es sospechosa o no en base a las transacciones del cliente.

2 Revisión de la literatura

2.1 Lavado de activos

No obstante, las definiciones encontradas para el lavado de activos [22],[23],[24],[25],[26],[27], y las variaciones que puede tener la expresión, todas se refieren a la intención de dar apariencia lícita a los recursos obtenidos ilegalmente, insertándose en la economía para uso posterior buscando ocultar su origen y propiedad. El lavado de activos es un conjunto de operaciones comerciales o financieras que buscan incorporar a la economía de cada país los recursos, bienes y servicios que se originan o están vinculados a actos ilegales [28]. Por otra parte, según la Interpol, el lavado de activos se define como cualquier acto o intento de ocultar o disfrazar la identidad de los ingresos obtenidos ilegalmente para que parezcan haberse originado de fuentes legítimas.

Los delitos económicos son aquellos que involucran instrumentos e instituciones del mercado financiero que buscan obtener ganancias a expensas de otros participantes del mercado [27]. Dichas actividades son ilegales porque violan las reglas de la economía de libre competencia, ya que los delincuentes realizan transacciones sin ningún sentido económico en el que la maximización de ganancias no sea su objetivo [27]. Es por eso que existe un régimen de financiamiento antilavado de activos y su objetivo principal es reducir las tasas de delincuencia profesional, la delincuencia organizada y el terrorismo y proteger a la sociedad en su conjunto según menciona Badal-Valero *et al.* [9]. El lavado de activos no es un acto aislado sino un proceso que implica una secuencia de actos concatenados en el tiempo y el espacio: la colocación, el ocultamiento y la integración de recursos [29], [30], [31]. Por otro lado, de acuerdo con Feng *et al.* [40], el anti lavado de activos se puede dividir en tres etapas: prevención, detección y sanciones.

Como resultado de estas actividades, [32] el HSBC fue multado por las autoridades estadounidenses a gran suma de \$ 1.9 mil millones (£ 1.2 mil millones) en un acuerdo sobre lavado de dinero, el mayor pagado en tal caso. Por lo tanto, los gobiernos, los reguladores financieros requieren que las instituciones financieras implementen procesos y procedimientos para prevenir o detectar el lavado de dinero, así como el financiamiento del terrorismo y otras actividades ilícitas en las que participan los lava-

dores de dinero. Por lo tanto, el antilavado de dinero (AML) es de importancia crítica para la estabilidad financiera nacional y la seguridad internacional. [33]

2.2 Machine learning

La generación de grandes volúmenes de datos hace que la aplicación de machine learning sea una realidad, pero qué es el machine learning, para Kang y Jamenson [11] machine learning es un conjunto de técnicas con las cuáles se puede obtener información útil de un conjunto de datos para acelerar el desarrollo de métodos de detección, diagnóstico y pronóstico de anomalías basados en datos. Por el contrario, Samuel [12], citado por Kang y Jamenson [11], define el machine learning como el campo de estudio que le da a las computadoras la capacidad de aprender sin ser programado explícitamente. Además, como se muestra en la figura 1, para Kang y Jamenson[11] el machine learning puede clasificarse en categorías de acuerdo a la forma de entrenamiento de los algoritmos, por ejemplo, si están entrenados mediante supervisión humana, se clasifican en aprendizaje supervisado, no supervisado, semi-supervisado y de refuerzo. En contraste si los algoritmos pueden aprender de forma incremental en la marcha pueden clasificarse como aprendizaje en línea o por lotes. Por otro lado, si funcionan comparando datos nuevos con datos conocidos, o si detectan patrones en los datos de entrenamiento y crean un modelo predictivo pueden clasificarse en aprendizaje basado en instancia o aprendizaje basado en modelos.

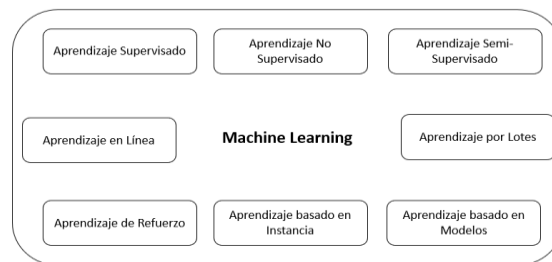


Fig. 1 Clasificación de machine learning según la forma de entrenamiento

2.3 Algoritmo de machine learning

Un algoritmo de aprendizaje automático es la combinación de una función de pérdida y una técnica de optimización. La función principal del algoritmo de aprendizaje automático, consiste en encontrar patrones en los datos para generar un modelo de aprendizaje automático que extraiga estos patrones y de acuerdo con Golmohammadi, Zaiane y Diaz [5] el reconocimiento de patrones en los datos utilizando algoritmos de aprendizaje supervisado tiene como objetivo detectar patrones que sean similares a las tendencias que se presentan en la actividad fraudulenta. Los algoritmos de aprendizaje automático pueden clasificarse en diferentes categorías que existen en el machine learning que de acuerdo con Kang y Jamenson[11], son aprendizaje supervisado, no supervisado, semi-supervisado y aprendizaje de refuerzo.

2.3.1 Algoritmos de aprendizaje supervisado

La característica principal de estos tipos de algoritmos de machine learning es que trabajan con datos que tienen etiquetada la variable o solución deseada de acuerdo con Kang y Jamenson[11]. Dentro de esta categoría de machine learning, la clasificación y la predicción son las tareas más representativas, y como ejemplo citamos el uso de los algoritmos de árboles de decisión, naive bayes, redes neuronales, SVM y KNN con la finalidad de clasificar muestras de manipulación de mercado utilizando un conjunto de datos estructurado por observaciones fraudulentas y normales, en el estudio realizado por Golmohammadi *et al.* [5]. En cuanto al algoritmo árboles de decisión, podemos afirmar que es de fácil interpretación, tienen un buen rendimiento y son escalables, pero presentan una desventaja que es el sobreajuste según afirma Golmohammadi *et al.* [5], pero según [13] y [14], citados por Golmohammadi *et al.* [5], este inconveniente puede solucionarse con los métodos de poda y ensamblaje como bosques aleatorios y árboles potenciados.

2.3.2 Algoritmos de aprendizaje no supervisado

Este tipo de algoritmos se caracteriza por trabajar con datos que no están etiquetados, por lo tanto, la búsqueda de patrones se organiza de una forma diferente mediante tareas de agrupación y reducción de la dimensionalidad. De acuerdo con Kang y Jamenson[11], para detectar anomalías en un conjunto de datos los algoritmos de agrupación son utilizados ampliamente para lograr este fin, y de esto tenemos una representación en la figura 2, elaborada por Amarasinghe *et al.* [15] en el cual se observa la distribución de los datos en un espacio bidimensional donde los grupos N1 y N2 representan los datos normales y O1 y O3 representan datos anómalos y de esto podemos entender que una anomalía es un dato muy distinto al resto. Básicamente este agrupamiento se puede formar mediante el uso de algoritmos como el K-means, algoritmos jerárquicos, isolation forest, entre otros.

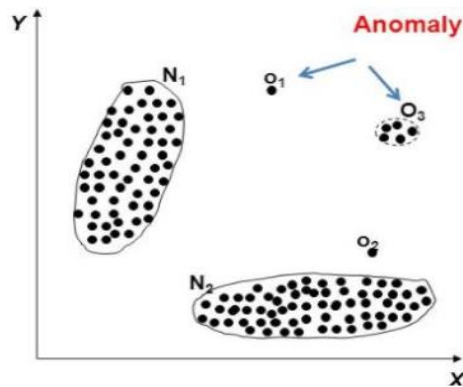


Fig. 2 Puntos de anomalías [15]

Otro aspecto respecto a esta categoría, según Brownlee [16], citado por Amarasinghe *et al.* [15], cuando los datos de entrenamiento no están disponibles es necesario usar algoritmos que pertenezcan a esta categoría o a la categoría de semi-supervisado.

2.3.3 Algoritmos de aprendizaje semi-supervisado

Definido por Kang y Jamenson [11] como una clase de tareas y técnicas de aprendizaje supervisado que utilizan datos no etiquetados en su mayoría (80% de 100%) para el entrenamiento. Además, el aprendizaje semi-supervisado se puede dividir en dos clases principales, aprendizaje inductivo y aprendizaje transductivo. Como ejemplo Tamil Selva y Wang [17], proponen el uso de las denominadas máquinas de Boltzmann restringidas como un enfoque de aprendizaje semi-supervisado que pueden usarse para el diagnóstico de la salud. Además estos algoritmos pueden trabajar con un conjunto de datos que tienen solo una minoría de datos etiquetados y obtener casi el mismo nivel de resultados. A la vez estos datos no etiquetados en la investigación de Xu *et al.* [18] se han dividido en categorías según la calidad que tienen los cuales pueden ser de mayor o menor influencia para mejorar la predicción.

2.3.4 Algoritmos de aprendizaje de refuerzo

Los algoritmos de aprendizaje de refuerzo tienen la finalidad de lograr que una máquina sea capaz de decidir mediante su propia experiencia, es decir una máquina entrenada con algoritmos de aprendizaje reforzado puede ser capaz de tomar decisiones, aunque no tenga conocimientos a priori. Esto lo logra mediante la extracción de acciones que han de ser elegidas en los diferentes estados con la finalidad de maximizar una recompensa, según menciona Sancho [19] al ejemplificar el proceso de decisión de Markov. En la figura 3 elaborada por Sutton y Barton, citada por Merino [20], se muestra la estructura básica del aprendizaje por refuerzo, se necesita de un agente en un estado determinado dentro de un medio ambiente. Este agente por cada acción que realice recibirá una determinada recompensa, y según obtenga las recompensas el agente podrá determinar qué acciones seguir para conseguir un determinado objetivo. En la mayoría de los casos este tipo de algoritmos de aprendizaje por refuerzo son empleados para la clasificación de secuencias de ADN, conducción de vehículos, en el apoyo de diagnósticos médicos, entre otros.

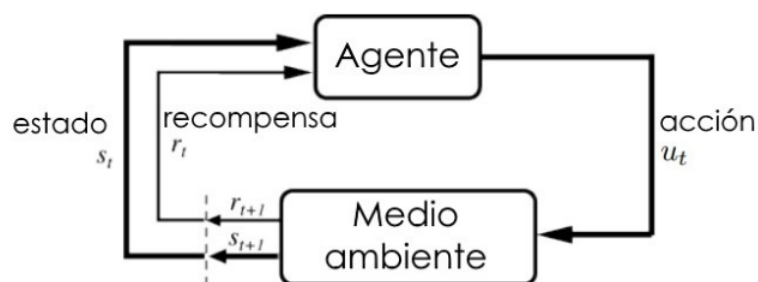


Fig. 3 Estructura básica del aprendizaje por refuerzo [20]

2.4 Modelo de aprendizaje automático

Un modelo de machine learning son algoritmos de machine learning que han sido entrenados de manera conjunta o unitaria para poder trabajar con nuevos datos de entrada con la finalidad de realizar la tarea para la cual fue programada, en muchos de los casos clasificación y agrupación. Los modelos de machine learning se pueden clasificar en varios grupos de acuerdo a su estructura y forma de trabajar según Noblejas [21]. Estos modelos tienen una fuerte base estadística debido a que trabajar con datos del mundo real hace casi imposible establecer un modelo preciso para realizar las predicciones.

3 Método de la revisión sistemática de la literatura

3.1 Necesidad de la revisión sistemática

La revisión sistemática de la literatura que se presenta en este estudio surge a partir de la necesidad de identificar qué modelos de aprendizaje automático se han utilizados para la detección de transacciones sospechosas de lavado de activos en las entidades financieras. En la tabla 1 se presenta la elaboración del objetivo de la investigación donde se resume el objeto de estudio y el propósito del estudio reforzado con los demás puntos.

Tabla 1. Elaboración del objetivo de la investigación

Campo	Valor
Objeto de estudio	Modelos de aprendizaje automático
Propósito	Identificar
Foco	Modelos - Algoritmos - Técnicas
Involucrados	Algoritmos de aprendizaje, Modelos de aprendizaje automático, Técnicas de aprendizaje automático, Aprendizaje automático
Factores de contexto	Bancos, Financieras, Agencias de Giros de dinero

3.2 Preguntas para la revisión sistemática

Para la definición y organización de las preguntas de investigación se tomó como referencia la finalidad de la investigación presentada en la sección anterior. En la tabla 2 se detallan las preguntas de investigación propuestas y la motivación para cada una de ellas con el objetivo de analizar la eficiencia de cada modelo de aprendizaje

automático que ha sido implementado, planteado y diseñado a la fecha para la detección de transacciones sospechosas de lavado de activos.

Adicionalmente en la tabla 3 se describe las preguntas bibliométricas propuestas para la evaluación de los diferentes artículos de investigación seleccionados para esta Revisión Sistemática de la Literatura de acuerdo a los factores de tiempo, tipo de artículo y el tema investigado.

Tabla 2. Preguntas de investigación y motivación

ID	Pregunta	Motivación
PI-1	¿Qué modelos de aprendizaje automático existen para la detección de transacciones sospechosas de lavado de activos?	Identificar los modelos de aprendizaje automático existentes para la detección de transacciones sospechosas de lavado de activos.
PI-2	¿Cuáles son los modelos de aprendizaje automático más utilizados para la detección de transacciones sospechosas de lavado de activos?	Analizar los modelos que comúnmente son utilizados para la detección de transacciones sospechosas de lavado de activos con la finalidad de evaluar el desempeño de cada uno de acuerdo a los distintos tipos de data analizada.
PI-3	¿Cuáles son los algoritmos de aprendizaje automático más utilizados para detección de transacciones sospechosas de lavado de activos?	Analizar los algoritmos que comúnmente se utilizan para la detección de transacciones sospechosas de lavado de activos con la finalidad de evaluar el desempeño de cada uno de acuerdo a los distintos tipos de data analizada.

Tabla 3. Preguntas Bibliométricas

ID	Pregunta	Motivación
PB-1	¿Cuál es la cantidad de publicaciones por tipo de fuente del artículo?	Determinar cuál es el tipo de fuente con mayores publicaciones respecto al tema de estudio

3.3 Definición de las cadenas de búsqueda

La estrategia PICO fue la seleccionada para la elaboración de la cadena de búsqueda a través de un proceso iterativo, en el cual se realizaron los ajustes convenientes para la selección de resultados.

Población:

Término principal 1: Modelos de aprendizaje automático

Términos alternos: Aprendizaje automático, algoritmos de aprendizaje.

Justificante: Identificar qué modelos existen dentro del aprendizaje automático para la detección de transacciones sospechosas de lavado de activos en entidades financieras.

Intervención:

Término principal 1: Transacciones sospechosas

Términos alternos: Transacciones fraudulentas, operaciones sospechosas, delito financiero, lavado de activos, lavado de dinero.

Justificante: Reducir el porcentaje de lavado de activos que se hacen a través de las entidades bancarias para mantener un equilibrio en la economía del país.

Resultado:

Término principal 1: Implementación

Términos alternos: diseño, clasificación, técnicas.

Justificante: Identificar los diferentes modelos de aprendizaje automático implementados y/o propuestos para la detección de transacciones sospechosas de lavado de activos en entidades financieras.

3.3.1 Idioma

El inglés fue el idioma elegido para definir la cadena de búsqueda, puesto que es el más utilizado para la elaboración de artículos en las librerías digitales seleccionadas.

Tabla 4. Términos en inglés y conectores lógicos a ser usados en la búsqueda

Concepto	Términos en inglés
Población	(model) and (machine learning or learning algorithm or automatic learning techniques)
Intervención	(suspicious transactions) and (fraudulent transactions or suspicious operations or money laundering)

Comparación	No aplica
Resultado	(focus or implementation or design or classification or techniques)
Contexto	No aplica

3.4 Criterios de inclusión y exclusion

Luego de ejecutar las respectivas cadenas de búsqueda en las diferentes librerías digitales, los resultados deben ser evaluados con la finalidad de definir cuáles son los estudios primarios que responden a las preguntas de investigación. Por esta razón, se definen los criterios de inclusión y exclusión para que cada artículo sea evaluado en la selección.

Criterios de inclusión

CI.1. Se consideran todos aquellos artículos provenientes de las siguientes librerías digitales indexadas IEEE Xplore Digital Library, ScienceDirect, ACM Digital Library y Springer Link.

CI.2. Los artículos deben provenir del área de aprendizaje automático, minería de datos y big data.

CI.3 Se aceptarán artículos que contengan estudios o análisis de algoritmos de aprendizaje automático para la detección de lavado de activos en entidades financieras.

CI.4. Se aceptarán artículos provenientes de revistas científicas y conferencias.

CI.5. Se considerarán todos los artículos que analicen los modelos de aplicado a la detección de transacciones sospechosas de lavado de activos en entidades financieras.

CI.6. Se considerará todos los artículos que contengan temas relacionados a la detección de transacciones sospechosas de lavado de activos en entidades financieras.

Criterios de exclusión

C.E.1. Serán excluidos los artículos duplicados.

C.E.2. No se considerará artículos que no estén publicados en inglés.

C.E.3. Serán rechazados los artículos de contenido similar, mediante un análisis de mayor impacto en la investigación.

C.E.4. Serán excluidos los artículos cuyo título no tenga relación con el objeto de estudio.

3.5 Criterios de calidad

A continuación, con la finalidad de evaluar la calidad de los estudios seleccionados se plantean tres criterios de calidad para evaluarlos. Dentro de la evaluación cada criterio tiene una escala de puntuación donde S=1, P=0.5 y N= 0. Los resultados que se obtengan serán presentados en la tabla 5.

Tabla 5. Criterios de evaluación de la calidad

N°	Criterio de evaluación de calidad
1	<p>¿La metodología utilizada para desarrollar el estudio se documentó apropiadamente?</p> <p>S: La metodología utilizada ha sido documentada apropiadamente. P: La metodología utilizada ha sido documentada parcialmente N:La metodología no ha sido documentado</p>
2	<p>¿Se ha documentado el alcance del estudio de forma clara?</p> <p>S:El alcance se ha documentado claramente. P:El alcance se ha documentado parcialmente N:No se ha documentado el alcance.</p>
3	<p>¿Los resultados obtenidos en el estudio sirvieron para responder a las preguntas de investigación?</p> <p>S:Los resultados del estudio ayudan a responder las preguntas de investigación de forma clara. P:Los resultados del estudio ayudan a responder parcialmente las preguntas de investigación. N:Los resultados del estudio no ayudan a responder las preguntas de investigación.</p>

3.6 Fuente de datos

Las bases de datos digitales consideradas por su relevancia científica para la selección de los artículos fueron:

- SCIENCE DIRECT (<https://www.sciencedirect.com/>)
- IEEE XPLORE Digital Library (<https://ieeexplore.ieee.org/>)
- ACM DIGITAL LIBRARY (<https://dl.acm.org/>)
- SPRINGER LINK(<https://link.springer.com>)

3.7 Procedimiento para la selección de estudios

Se considera el siguiente procedimiento para la selección de los estudios:

- Paso 1: se procedió a ejecutar la cadena de búsqueda PICO, en las bases de datos seleccionadas. A los resultados se aplicó los criterios de inclusión y exclusión conforme a lo descrito en la tabla 6. Las referencias de los artículos que se obtuvo como resultado fueron guardadas para su posterior refinamiento.
- Paso 2: se revisaron los contenidos de los artículos resultantes de la ejecución del Paso 1 excluyendo los artículos duplicados y de contenidos similar, dando relevancia a los que tengan el contenido más completo con relación al objeto de estudio.
- Paso 3: se revisaron los resúmenes de los artículos previamente seleccionados en el Paso 2 para proceder con la selección artículos que se encuentren dentro de los últimos 10 años, que contengan estudios o análisis de algoritmos de aprendizaje automático para la detección de anomalías con relación al objeto de estudio.
- Paso 4: se procedió con la realización de una revisión preliminar del contenido de los artículos seleccionados luego del Paso 3, con especial atención a los estudios que analicen modelos de predicción aplicado a la detección de transacciones sospechosas en entidades financieras.

Tabla 6. Procedimientos y criterios de inclusión

Procedimiento	Criterio de selección
Paso 1	C.I.1 – C.I.2 – C.E.2 – C.E.5
Paso 2	C.E.1 – C.I.3
Paso 3	C.E.3 – C.I.4
Paso 4	C.I.5 – C.I.6

4 Resultados

4.1 Resultados de la búsqueda

De acuerdo con los pasos definidos en la sección anterior, el primer paso para la selección de estudios consiste en la ejecución de las cadenas de búsqueda en las libre-

rías digitales elegida para este estudio. En la tabla 7 se muestran los resultados, la fecha en la que fue ejecutada la cadena de búsqueda y las cadenas de búsqueda.

Tabla 7. Resultados de búsqueda

Cadena de búsqueda		
Base de Datos	Fecha	Total
SCIENCE DIRECT	Octubre 2019	13
("suspicious transactions" OR "money laundering") AND machine learning OR ("anti-money laundering" OR "money laundering") AND machine learning Article Type: Review Articles, Research Articles		
IEEE Xplore Digital Library	Junio 2019	45
"All Metadata": machine learning AND suspicious transactions OR machine learning AND money laundering OR (anti-money laundering OR money laundering) AND machine learning OR (suspicious transactions OR money laundering) AND machine learning OR (machine learning algorithm OR machine learning model) AND money laundering OR (machine learning algorithm OR machine learning model) AND suspicious transactions		
SPRINGER Link	Octubre 2019	114
machine learning AND model AND algorithm AND suspicious AND transactions AND money AND laundering		
ACM Digital Library	Octubre 2019	313
acmdlTitle:("money laundering" "suspicious transactions" "anti-money laundering" "machine learning") AND recordAbstract:("anti-money laundering" "money laundering" "suspicious transactions" "machine learning model" "machine learning algorithm") AND content.ftsec:("anti-money laundering" "money laundering" "suspicious transactions" "machine learning model" "machine learning algorithm") Content Formats: PDF		

4.2 Resultados de filtros aplicados

4.2.1 Selección de estudios primarios

Para la selección de los estudios primarios, los resultados de la búsqueda se sometieron a los siguientes pasos:

Paso 1: El conjunto de artículos resultantes tras la ejecución de las cadenas de búsqueda fueron sometidos a los criterios de inclusión y exclusión definidos en este paso. Para ello nos basamos en la lectura de los títulos y resumen con el fin de acotar el tiempo empleado en la lectura de artículos que no tienen relación con nuestro objeto de estudio. Para los resultados de este paso fue necesario identificar los links de acceso y almacenarlos en un archivo excel, bajo la estructura de una tabla.

Paso 2: La lista de artículos resultantes del paso 1, se sometieron a una revisión más profunda aplicando los criterios de inclusión y exclusión definidos en este paso, como la duplicidad de artículos. En este caso un artículo de ACM coincidió con otro de IEEE Xplore Digital Library. Los artículos que no pasaron este paso fueron tachados en la tabla y los que pasaron se usarán como base en el paso 3.

Paso 3: Los resultados del paso 2 se revisaron en base a los criterios definidos para este paso, para esto fue necesario realizar una lectura de las conclusiones y la metodología además del resumen. Para esto fue necesario descargar los artículos.

Paso 4: Teniendo en cuenta los criterios de este paso se procedió a revisar el contenido total de los artículos resultantes del paso anterior. En este paso se procedió de forma más cautelosa al elegir los artículos, puesto que los resultados de este paso nos ayudarán a resolver las preguntas de investigación y con ello lograr nuestro objetivo en este estudio.

En la tabla 8 se muestran los resultados de la selección de los artículos de estudio primario resultante.

Tabla 8. Resultados del proceso de selección de estudios primarios

Base de datos	Artículos descubiertos	Paso 1	Paso 2	Paso 3	Paso 4
SCIENCE DIRECT	13	7	4	4	3
IEEE Xplore Digital Library	45	16	14	13	7

ACM Digital Library	313	6	6	6	6
SPRINGER Link	114	8	7	6	4
Total	485	37	31	29	20

4.2.2 Evaluar Calidad de los estudios

Tabla 9. Resultados de la evaluación de la calidad

Nombre Artículo	C1	C2	C3	Total
Finding shell company accounts using anomaly detection	1	0.5	1	2.5
A scan statistics based suspicious transactions detection model for Anti-Money Laundering (AML) in financial institutions	1	0.5	1	2.5
Application of cluster-based local outlier factor algorithm in anti-money laundering	1	1	1	3
Developing an intelligent data discriminating system of anti-money laundering based on SVM	1	1	1	3
Detecting Stock Market Manipulation using Supervised Learning Algorithms	0.5	0	1	1.5
Finding suspicious activities in financial transactions and distributed ledgers	1	0.5	1	2.5
Fraud Risk Monitoring System for E-Banking Transactions	1	0.5	1	2.5
Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods	0.5	0.5	0.5	1.5
Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case	0.5	0.5	0.5	1.5
Clustering based anomalous transaction reporting	1	0	0.5	1.5
Machine Learning : Fundamentals 4 . 1 Types of Machine Learning	1	0	0.5	1.5
Towards a New Data Mining-Based Approach for Anti-Money Laundering in an International Investment Bank	1	0.5	0.5	1.5
Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review	1	0.5	0.5	1.5

4.3 Análisis bibliométrico (E. Análisis bibliométrico)

En esta parte del documento se desarrolla el análisis de la tendencia de los artículos encontrados para el presente estudio.

A. ¿Cuál es la cantidad de publicaciones por tipo de artículo?

En la Fig. 6 se muestra la cantidad de publicaciones por tipo de artículo. Podemos observar que los artículos de conferencia (Conference Proceedings) representan el 77% del total de artículos seleccionados para esta revisión sistemática de la literatura; consecutivamente tenemos a los artículos en revista (Journal Article) con un 23%. De este análisis se puede concluir que los artículos de conferencia son la mayor fuente de estudios sobre modelos de aprendizaje automático

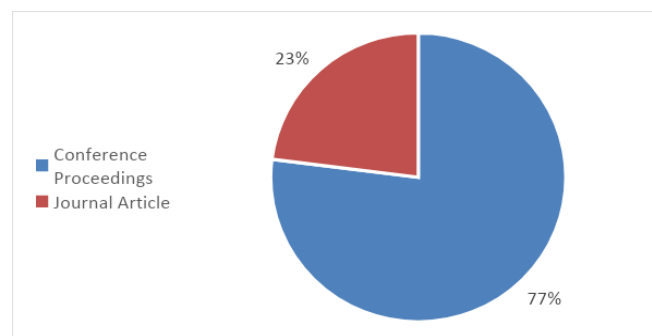


Fig. 4 Cantidad de publicaciones por tipo

Preguntas de Investigación

En esta parte del documento se desarrolla el análisis de la tendencia de los artículos encontrados para el presente estudio a través de una revisión sistemática de la literatura. Este análisis se hace de acuerdo con los factores como tipo de artículo.

A. ¿Qué modelos de aprendizaje automático existen para la detección de transacciones sospechosas de lavado de activos?

Al realizar el análisis de la información de los diferentes artículos previamente seleccionados se encontraron diferentes modelos aplicados a la detección de transacciones sospechosas de lavado de activos, cada uno tratando de abordar al máximo la naturaleza de la actividad ilícita.

Modelo de Zengan:

Zengan [3] para su modelo se basa en la inteligencia artificial, la máquina de vectores de soporte, la detección de valores atípicos y el análisis del punto de ruptura los cuales son aplicados en las instituciones financieras para mejorar la capacidad de

detección de transacciones sospechosas en el procesamiento de los millones de datos. Partiendo desde este enfoque Zengan [3] diseña el algoritmo factor de valor atípico local basado en agrupaciones (CBLOF). Con la finalidad de ayudar a las instituciones financieras en la detección de patrones de comportamiento transaccional sospechosos de lavado de activos, este algoritmo combina la agrupación no supervisada basada en la distancia con la detección local de valores atípicos. Esta combinación hace que el algoritmo presente una alta velocidad de procesamiento y una precisión satisfactoria en la identificación de transacciones sospechosas de lavado de activos de acuerdo con la prueba experimental que realizó. Dentro de las ventajas del modelo propuesto por Zengan [3], se destaca que este modelo no hace uso de datos para el entrenamiento ni tampoco requiere que se le asigne un número de grupos, con lo cual se puede dar solución al inconveniente de falta de datos para trabajos similares. Otro aspecto a resaltar del modelo de Zengan [3] es su autoadaptabilidad a la evolución de los métodos de lavado de activos debido a que tiene la capacidad de reconocer nuevos patrones de transacciones sospechosas.

Modelo de Tang y Yin:

Por otro parte Tang y Yin [4] con el fin de determinar el comportamiento del cliente debido a que estos pueden usar entidades fantasmas para pasar desapercibidos ante los mecanismos de detección de transacciones sospechosas, crea un modelo basado en las máquinas de vectores de soporte (SVM). De acuerdo con los autores, existen ciertas características esenciales para construir un perfil de comportamiento del cliente que son: suma de transacciones, frecuencia de transacciones, ciclo comercial, alternancia de tipos de negocios y cambio de socios que cooperan.

El modelo se diseña en base a una función de kernel (RBF) basado en la definición de distancia (HVDM) que extiende el algoritmo de C-SVM supervisado y el algoritmo SVM de una clase no supervisada. Con la finalidad de probar la efectividad de su modelo, Tang y Yin [4] realizaron un experimento con datos del Banco de Agricultura de Wuhan el cual tenía un registro de 5000 cuentas, 1,2 millones de registros durante 7 meses. Estos datos fueron mezclados con 80 cuentas inusuales simuladas cuyas características se desviaban de los grupos pares. En la estructura de datos utilizados para esta investigación, el id del cliente contiene algunas características más del cliente como la suma de capital registrado, el tipo de negocio principal, etc. Además también se considera la fecha, id de transferencia.

Los resultados obtenidos se muestran en la tabla 10, donde DR representa la tasa de detección, FPR la tasa de falsos positivos, C parámetro de castigo de clasificación incorrecta y g el factor de control.

Tabla 10. Resultados de la detección inusual

C	g	DR	FPR
50	0.5	63.27%	6.8%
100	1	69.13%	5.4%
200	2	67.8%	3.4%

En base a los resultados obtenidos Tang y Yin concluyen que su modelo es eficiente para la detección de comportamientos sospechosos en base al comportamiento del cliente. A la vez este modelo tiene un mejor rendimiento en el manejo de los datos y en la precisión al momento de la detección.

Modelo de Liu y Zhang:

Por el contrario, Liu y Zhang [2] propone un modelo basado en la estadística de escaneo, el cual tiene la finalidad de detectar secuencias sospechosas en las transacciones financieras. Para poder trabajar con este modelo, es necesario transferir el problema de detección de actividades sospechosas a un problema de estadística de escaneo con la finalidad de poder definir el flujo de trabajo. Para lograr esta transferencia es necesario realizar los siguientes pasos propuestos por [2].

- Recopilar datos de la institución financiera como información de los cliente, cuentas y transacciones

Tabla 11. Estructura de los datos recopilados

Información	Detalles
Cliente	Tipo de cliente (natural/jurídico), Escala de la corporación, nivel de riesgo individual, entre otros
Cuenta	Información numérica, dirección del flujo, tipo de operación, la moneda, la fuente, el destino, entre otros.
Transacciones	Hora de la transacción, fecha de la transacción, monto de la transacción, entre otros.

- Procesar los datos y asignar un nivel de riesgo para calcular el umbral de monto de la transacción que determinará el evento inusual para que el sensor pueda escanear y devolver las transacciones sospechosas

- Determinar las transacciones que tienen una alta probabilidad de ser nombradas como sospechosas

Al probar el modelo propuesto por [2], se utilizaron los datos de un banco comercial de Shanghái. Se procedió de la siguiente manera para el entrenamiento.

En primer lugar, se recopilaron y clasificaron los datos de las instituciones financieras para posteriormente tratarlas y dividir las en dos categorías, transacciones normales y sospechosas.

En segundo lugar, de los datos recopilados mediante una selección aleatoria se separa una cantidad determinada de datos normales y sospechosos con el objetivo de calcular el umbral de monto de transacción que determinará cuando un evento es inusual. Para este caso se seleccionaron las transacciones de 640 cuentas que están etiquetadas como normales y 35 cuentas cuyas transacciones representan las transacciones sospechosas.

En tercer lugar, utilizan métricas de rendimiento de clasificación estándar, sensibilidad / recuerdo y especificidad para determinar si las transacciones tienen una alta probabilidad de ser declaradas como sospechosas o normales.

Por último, se aplica el modelo de estadística de escaneo para evaluar las secuencias de las transacciones con la finalidad de determinar aquellas transacciones que están por encima del umbral establecido. Para el caso de estudio se estableció 0.8 como la probabilidad de los senderos de Bernoulli y 5 días como la ventana de escaneo.

Los resultados de la ejecución del algoritmo están representados en la tabla 12 donde podemos observar que la sensibilidad tiene un valor de 0.516 lo cual indica que es necesario ajustar un poco más ese aspecto.

Tabla 12. Resultados de la ejecución

P Value	Sensibilidad	Especificidad
0.8	0.516	0.949

Modelo de Luna, Palshikar, Apte, y Bhattacharya:

Luna *et al.* [1] tiene como objetivo principal identificar compañías fantasmas que son utilizados como vía para el lavado de activos. Plantean identificar estas compañías basándose solamente en las transacciones bancarias que estas pueden hacer como parte de su supuesta labor cotidiana. Para este fin utilizan algoritmos de detección de

anomalías los cuáles serán sometidos a encontrar patrones anormales en los datos de las transacciones tanto de las compañías honestas como de las fantasmas. Para probar su modelo realizan un experimento con una data de 10 000 compañías honestas y 60 compañías fantasmas al cual aplican tres algoritmos de detección de anomalías, el primero el RRS (Reliable Route Selection) el cual obtuvo una precisión de 0.16, el segundo fue el FastVOA el cual obtuvo una precisión de 0.49 y por último aplicaron el LOF (Local Outlier Factor) donde la precisión fue mayor al obtenerse un 0.98. Según consideración de Luna *et al.*[1], el LOF es el algoritmo más simple de usar en la práctica ya que se puede ajustar fácilmente su umbral.

Modelo de Larik y Haider:

Sin embargo, en el estudio realizado por Larik y Haider [10] se presenta un modelo que sugiere un nuevo algoritmo de agrupamiento denominado TEART (versión modificada de EART) y un índice de anomalía llamado AICAF. A su vez el modelo hace uso del Análisis de Componentes Principales (PCA) para la reducción de la dimensionalidad juntamente con K-means con el fin de obtener una estimación inicial del número de grupos presentes en el conjunto de datos. Por otro lado, K-means se utiliza en este modelo para comparar el rendimiento de TEART. El modelo en primer lugar realiza el agrupamiento mediante los algoritmos K-means y TEART donde se calcula y se guarda la distancia de cada registro de cliente desde su promedio de grupo. Luego se realiza el cálculo de anomalías a través del índice de anomalías denominado AICAF, el cuál mide la desviación del monto de la transacción y la frecuencia de tipos similares de transacciones del comportamiento establecido del clúster al que pertenece el cliente.

- B.** ¿Cuáles son los modelos de aprendizaje automático más utilizados para la detección de transacciones sospechosas de lavado de activos?

Partiendo desde el concepto de modelo de aprendizaje automático en esta investigación y en base a la revisión de los diferentes estudios; se puede concluir que no existe un modelo con la misma estructura que haya sido utilizada más de una vez, sino que se realiza una combinación de diferentes algoritmos y técnicas cada uno de estos enfocándose en abarcar un mayor campo de acción para abordar el problema en su totalidad. Por otro lado si se puede afirmar que los algoritmos de aprendizaje no supervisado son los que más se han utilizado en los diferentes estudios, debido a que los datos etiquetados para esta problemática son escasos, y a la misma vez estos pueden dar falsos positivos.

- C.** ¿Cuáles son los algoritmos de aprendizaje automático más utilizados para detección de transacciones sospechosas de lavado de activos?

Mediante el estudio se ha podido identificar que los algoritmos de agrupación son los más usados en los modelos de machine learning. Podemos mencionar que el algorit-

mo SVM es uno de los que se utilizan más en los diferentes estudios planteados en las diferentes publicaciones. De acuerdo con Tang y Yin [4], el algoritmo SVM es adecuado para el diseño de clasificadores y el descubrimiento de comportamiento inusual entre conjuntos de datos heterogéneos de alta dimensionalidad.

También Liu y Zhang [2] menciona a Tang y Yin [4] haciendo uso del algoritmo para tratar con datos de transacciones etiquetadas para descubrir actividades sospechosas de lavado de activos con la finalidad de seleccionar las cuentas o clientes sospechosos. Además en el estudio realizado por Hajek y Henriques [8] plantea el uso del algoritmo SVM, en este caso el SVM no lineal.

Por otra parte las variantes del algoritmo de árboles de decisión están también dentro de los más usados. La mayor ventaja que tiene este algoritmo es su capacidad de interpretación de las reglas generadas a partir del modelo según Hajek y Henriques [8]. En el estudio realizado por Guo *et al.* [7], se construye el algoritmo denominado clasificador de bosque aleatorio paralelo, el cual en su estructura básica contiene múltiples árboles de decisión que son entrenados por separado del muestreo de arranque múltiple del conjunto de entrenamiento con el fin de determinar la representación de la clase a la que pertenece, si es de la clase de fraude o genuino.

Otro caso de aplicación del algoritmo árbol de decisión y random forest se presenta en el estudio realizado por Badal-Valero *et al.* [9] con la finalidad de apoyar en el análisis de la base de datos de operaciones de un macro-caso sobre lavado de activos orquestado entre una empresa principal y un grupo de proveedores.

5 Conclusiones

En el presente estudio se han analizado los artículos resultantes luego de aplicar la metodología de la revisión sistemática de la literatura, los cuales fueron 20 artículos que fueron seleccionados de las diferentes librerías digitales indexadas de gran relevancia. Del mismo modo en el análisis bibliométrico se estudia la cantidad de publicaciones por tipo de artículo, en el cual el 77% de los artículos son de tipo journal articles o artículos de revisión y el 23% son artículos de conferencias. Por otro lado, el estudio tiene como objetivo identificar métodos de aprendizaje automático implementados, diseño o propuestos para la detección de transacciones sospechosas de lavado de activos en entidades financieras a través del análisis de los diferentes artículos seleccionados, y estos dieron como resultados 5 métodos de aprendizaje automático empleados cada uno tratando de abarcar el mayor campo de acción del delito. En algunos casos aún se necesita mejorar la efectividad en la precisión de detección puesto que para ser confiables se requiere cierto grado de exactitud, mayormente hablamos de un 90% a 95%, pero esto se prevé que se logrará a medida que se vaya ajustando el algoritmo con el entrenamiento de nuevos datos. Cabe recalcar que la mayoría de los modelos están basados en algoritmos de clustering como punto inicial, debido a que se necesita realizar cierta agrupación de clientes, cuentas u otras características desde la cual se toma como punto inicial para determinar si es sospechoso de lavado de activos.

Referencias

1. D. K. Luna, G. K. Palshikar, M. Apte, and A. Bhattacharya, "Finding shell company accounts using anomaly detection," in *ACM International Conference Proceeding Series*, 2018, pp. 167–174.
2. X. Liu and P. Zhang, "A scan statistics based suspicious transactions detection model for Anti-Money Laundering (AML) in financial institutions," in *Proceedings - 2010 International Conference on Multimedia Communications, Mediacom 2010*, 2010, pp. 210–213.
3. G. Zengan, "Application of cluster-based local outlier factor algorithm in anti-money laundering," in *Proceedings - International Conference on Management and Service Science, MASS 2009*, 2009.
4. J. Tang and J. Yin, "Developing an intelligent data discriminating system of anti-money laundering based on SVM," in *2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005*, 2005, pp. 3453–3457.
5. K. Golmohammadi, O. R. Zaiane, and D. Díaz, "Detecting Stock Market Manipulation using Supervised Learning Algorithms."
6. R. D. Camino, R. State, L. Montero, and P. Valtchev, "Finding suspicious activities in financial transactions and distributed ledgers," in *IEEE International Conference on Data Mining Workshops, ICDMW, 2017*, vol. 2017-November, pp. 787–796.
7. C. Guo, H. Wang, H.-N. Dai, S. Cheng, and T. Wang, "Fraud Risk Monitoring System for E-Banking Transactions," in *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 100–105.
8. P. Hajek and R. Henriques, "Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods," *Knowledge-Based Syst.*, vol. 128, pp. 139–152, Jul. 2017.
9. E. Badal-Valero, J. A. Alvarez-Jareño, and J. M. Pavía, "Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case," *Forensic Sci. Int.*, vol. 282, pp. 24–34, Jan. 2018.
10. A. S. Larik and S. Haider, "Clustering based anomalous transaction reporting," in *Procedia Computer Science*, 2011, vol. 3, pp. 606–610.
11. M. Kang and N. J. Jameson, "Machine Learning : Fundamentals 4 . 1 Types of Machine Learning," pp. 85–109, 2018.
12. S. Arthur, *Some Studies in Machine Learning Using the Game of Checkers*, vol. 3, no. 4. 1959.
13. L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
14. R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods," *Ann. Stat.*, vol. 26, no. 5, pp. 1651–1686, Oct. 1998.
15. T. Amarasinghe, A. Aponso, and N. Krishnarajah, "Critical analysis of machine learning based approaches for fraud detection in financial transactions," in *ACM International Conference Proceeding Series*, 2018, pp. 12–17.
16. Trading Economics, "Sri Lanka GDP Growth Rate," 2017. [Online]. Available: tradingeconomics.com/sri-lanka/gdpgrowth. [Accessed: 10-Oct-2017].
17. Tamilselvan, P. and Wang, P., "Failure diagnosis using deep belief learning based health state classification". *Reliability Engineering & System Safety* 115: 124–135.
18. Z. Xu, I. King, and M. R. Lyu, *More Than Semi-supervised Learning: A unified view on Learning with La-beled and Unlabeled Data*. LAP LAMBERT Academic Publishing,

2010. [Online]. Available: <http://www.amazon.com/More-Than-Semi-supervised-Learning-Unlabeled/dp/3843379106>.
19. F. Sancho, “Aprendizaje por refuerzo: algoritmo Q Learning”, Fernando Sancho Caparini, 2019. [En línea]. Disponible en: <http://www.cs.us.es/~fsancho/?e=109>. [Accedido: 21-oct-2019]
 20. M. Merino, “Conceptos de inteligencia artificial: qué es el aprendizaje por refuerzo”, 27-ene-2019. [Comentario en una entrada en un blog]. Disponible en: <https://www.xataka.com/inteligencia-artificial/conceptos-inteligencia-artificial-que-aprendizaje-refuerzo>. [Accedido: 20-oct-2019]
 21. R. Noblejas, “Estudio de algoritmos de detección de anomalías y propuesta de recomendaciones para su aplicación a entornos de ciberseguridad,” Universidad Politécnica de Madrid, 2016.
 22. N. A. Le Khac, S. Markos, M. O’Neill, A. Brabazon, and M.-T. Kechadi. An investigation into Data Mining approaches for Anti Money Laundering. In Proceedings of International Conference on Computer Engineering and Applications (ICCEA 2009), 2009.
 23. S. N. Wang and J. G. Yang. A Money Laundering Risk Evaluation Method Based on Decision Tree. In 2007 International Conference on Machine Learning and Cybernetics, volume 1, pages 283–286, Aug. 2007.
 24. L.-T. Lv, N. Ji, and J.-L. Zhang. A RBF neural network model for anti-money laundering. In 2008 International Conference on Wavelet Analysis and Pattern Recognition, volume 1, pages 209–215, Aug. 2008.
 25. Y. Wang, D. Xu, H. Wang, K. Ye, and S. Gao. Agent-oriented ontology for monitoring and detecting money laundering process. In Proceedings of the 2nd international conference on Scalable information systems, page 81. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.
 26. P. Umadevi and E. Divya. Money laundering detection using TFA system. In International Conference on Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012), pages 1–8, Dec. 2012.
 27. R. Dreżewski, J. Sepielak, and W. Filipkowski. System supporting money laundering detection. *Digital Investigation*, 9(1):8–21, June 2012.
 28. C. d. C. d. A. F. COAF. Coaf. lavagem de dinheiro: um problema mundial. Brasília/DF, 1999.
 29. P. Umadevi and E. Divya. Money laundering detection using TFA system. In International Conference on Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012), pages 1–8, Dec. 2012.
 30. A. S. A. d. M. Pitombo. Lavagem de dinheiro: a tipicidade do crime antecedente. Ed. Revista dos Tribunais, 2003
 31. T.-M. Cheong and Y.-W. Si. Event-based approach to money laundering data analysis and visualization. In Proceedings of the 3rd International Symposium on Visual Information Communication, page 21. ACM, 2010.
 32. Z. Chen, L. D. Van Khoa, E. N. Teoh, A. Nazir, E. K. Karuppiah, and K. S. Lam, “Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review,” *Knowl. Inf. Syst.*, vol. 57, no. 2, pp. 245–285, 2018.
 33. N. A. Le-Khac, S. Markos, and M. T. Kechadi, “Towards a new data mining-based approach for anti-money laundering in an international investment bank,” *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 31 LNICST, pp. 77–84, 2010.

APÉNDICE

Artículos seleccionados

ID	Biblioteca	Título	Autor	Año	Tipo Documento
1	IEEE Xplore Digital Library	A scan statistics based suspicious transactions detection model for Anti-Money Laundering (AML) in financial institutions	X. Liu and P. Zhang	2010	Conference Paper
2	IEEE Xplore Digital Library	Application of cluster-based local outlier factor algorithm in anti-money laundering	G. Zengan	2009	Conference Paper
3	IEEE Xplore Digital Library	Developing an intelligent data discriminating system of anti-money laundering based on SVM	J. Tang and J. Yin	2005	Conference Paper
4	IEEE Xplore Digital Library	Detecting Stock Market Manipulation using Supervised Learning Algorithms	K. Golmohammadi, O. R. Zaiane, and D. Díaz	2014	Conference Paper
5	IEEE Xplore Digital Library	Finding suspicious activities in financial transactions and distributed ledgers	R. D. Camino, R. State, L. Montero, and P. Valtchev	2017	Conference Paper
6	IEEE Xplore Digital Library	Fraud Risk Monitoring System for E-Banking Transactions	C. Guo, H. Wang, H.-N. Dai, S. Cheng, and T. Wang	2018	Conference Paper
7	Science Direct	Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning	P. Hajek and R. Henriques	2017	Journal Article

		ning methods			
8	Science Direct	Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case,	E. Badal-Valero, J. A. Alvarez-Jareño, and J. M. Pavía	2018	Journal Article
9	Science Direct	Clustering based anomalous transaction reporting	A. S. Larik and S. Haider	2011	Journal Article
10	IEEE Xplore Digital Library	Machine Learning : Fundamentals 4 . 1 Types of Machine Learning	M. Kang and N. J. Jameson	2018	Journal Article
11	Springer	Web Information Systems and Applications	Weiwei Ni · Xin Wang · Wei Song · Yukun Li	2019	Conference Paper
12	Springer	Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review	Zhiyuan Chen · Le Dinh Van Khoa1, Amril Nazir, Ettikan Kandasamy Karuppiyah, · Kim Sim Lam	2017	Journal Article
13	Springer	Applying Data Mining in Money Laundering Detection for the Vietnamese Banking Industry	Dang Khoa Cao, Phuc Do	2012	Journal Article
14	Springer	Towards a New Data Mining-Based Approach for Anti-Money Laundering in an International Investment Bank	Le-Khac, Nhien An Markos, Sammer Kechadi, Mohand Tahar	2010	Journal Article
15	ACM Digi-	Finding shell com-	Luna, Devendra	2018	Conference

	tal Library	pany accounts using anomaly detection	Kumar Palshikar, Girish Keshav Apte, Manoj Bhattacharya, Arnab		Paper
16	ACM Digital Library	Comparing Data Mining Techniques for Anti-Money Laundering Aplicações de Técnicas de Aprendizagem de Máquina no Combate à Lavagem de Dinheiro	Fernando, Luis Dias, Carvalho Parreiras, Fernando Silva	2019	Journal Article
17	ACM Digital Library	Critical analysis of machine learning based approaches for fraud detection in financial transactions	Amarasinghe, Thushara Aponso, Achala Krishnarajah, Naomi	2018	Conference Paper
18	ACM Digital Library	Event-based approach to money laundering data analysis and visualization	Cheong, Tat-Man Si, Yain-Whar	2010	Conference Paper
19	ACM Digital Library	Information extraction of regulatory enforcement actions: From anti-money laundering compliance to countering terrorism finance	Plachouras, Vassilis Leidner, Jochen L.	2015	Conference Paper
20	ACM Digital Library	Research on the Periodical Behavior Discovery of Funds in Anti-money Laundering Investigation	He, Shiliang Qu, Zhenxin	2019	Journal Article