Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos del Sector Financiero Popular y Solidario del Ecuador

Chamba Macas Jonathan Xavier

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

Msc. Diaz Zúñiga, Magi Paúl

07 de febrero del 2023

Reporte de verificación de cumplimiento

Informe de originalidad

NOMBRE DEL CURSO
Revisión Tesis MGSI

NOMBRE DEL ALUMNO
JONATHAN XAVIER CHAMBA MACAS

NOMBRE DEL ARCHIVO
JONATHAN XAVIER CHAMBA MACAS - Documento sin título

SE HA CREADO EL INFORME
26 ene 2023

Resumen

Fragmentos marcados	6	2 %
Fragmentos citados o entrecomillados	0	0 %

Coincidencias de la Web

vlex.ec	2	0,9 %
worldbank.org	2	0,7 %
facebook.com	1	0,4 %
powerdata.es	1	0,3 %

Firma:

MAGI PAUL DIAZ ZUNIGA
Firmado digitalmente por MAGI PAUL DIAZ ZUNIGA
Fecha: 2023.02.13 09:03:14 -05'00'

Diaz Zúñiga, Magi Paúl

Director

C.C.: 1707249072

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Certificación

Certifico que el trabajo de titulación, "**Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos del Sector Financiero Popular y Solidario del Ecuador**" fue realizado por el señor **Chamba Macas Jonathan Xavier** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 07 de febrero de 2023

Firma

MAGI PAUL DIAZ ZUNIGA

Firmado digitalmente por MAGI PAUL DIAZ ZUNIGA
Fecha: 2023.02.13 09:04:47 -05'00'

Diaz Zúñiga, Magi Paúl

Director

C.C.: 1707249072

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Responsabilidad de autoría

Yo **Chamba Macas Jonathan Xavier**, con cédula de ciudadanía n° 1717928681, declaro que el contenido, ideas y criterios del trabajo de titulación***:*** **Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos del Sector Financiero Popular y Solidario del Ecuador** es de mí autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 07 de febrero de 2023

Firma

Chamba Macas Jonathan Xavier

C.C.: 1717928681

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Autorización de publicación

Yo, **Chamba Macas Jonathan Xavier**, con cédula de ciudadanía n° 1717928681, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos del Sector Financiero Popular y Solidario del Ecuador** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 07 de febrero de 2023

Firma

Chamba Macas Jonathan Xavier

C.C.: 1717928681

Dedicatoria

Este trabajo va dedicado a mi familia, en especial a mis abuelitos Ángel y Juana, a mi mami Bélgica y a mi ñaña Sofía.

Agradecimiento

Agradezco a Dios, a mi familia y a todos quienes forman parte de la Universidad de la Fuerzas Armadas ESPE.

Índice de contenidos

Índice de figuras

Índice de Tablas

Resumen

De acuerdo con el índice de vulnerabilidad ante lavado de activos publicado por el Instituto de Gobernanza de Basilea, al 2019, el Ecuador se ubicó en el puesto 29 de un total de 125 países analizados, es decir, entre los países más proclives a verse afectados por prácticas de lavado de activos, siendo el sistema financiero el principal afectado. En el caso de Ecuador, el sistema financiero está conformado por las instituciones financieras del sector público, privado y popular y solidario, en donde el sector financiero popular y solidario representa aproximadamente un tercio de las captaciones del sistema financiero nacional.

Con esta premisa, el presente trabajo tuvo como objetivo desarrollar una propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósito en la información que es reportada por las instituciones del sector financiero popular y solidario al Organismo de Control, y de esta manera brindar una herramienta de apoyo para el análisis de potenciales casos de lavado de dinero. Para este fin, las actividades ejecutadas se enmarcaron en la metodología de descubrimiento de conocimiento en bases de datos (KDD por sus siglas en inglés), la cual contempló desde la comprensión del caso de estudio hasta la aplicación de técnicas de minería de datos, y la presentación y evaluación de los resultados obtenidos a través de una herramienta de inteligencia de negocios.

Palabras clave: identificación de anomalías, minería de datos, extracción de conocimiento, inteligencia de negocios, lavado de dinero.

Abstract

According to the asset laundering vulnerability index published by the Basel Institute of Governance, as of 2019, Ecuador ranked 29th out of a total of 125 countries analyzed, that is, among the countries most likely to be affected for money laundering practices, with the financial system being the main affected. In the case of Ecuador, the financial system is made up of financial institutions from the public, private and popular and solidarity sectors, where the popular and solidarity financial sector represents approximately a third of the deposits of the national financial system.

With this premise, the present work aimed to develop a model proposal for the identification of anomalies in the balances of deposit accounts in the information that is reported by the institutions of the popular and solidarity financial sector to the Supervisory Entity and from this way to provide a support tool for the analysis of potential money laundering cases. For this purpose, the activities carried out were framed in the knowledge discovery methodology in databases (KDD for its acronym in English), which contemplated from the understanding of the case study to the application of data mining techniques, and the presentation and evaluation of the results obtained through a business intelligence tool.

Keywords: anomaly detection, data mining, knowledge discovery in databases, business intelligence, money laundering.

Capítulo I

El problema de investigación

Antecedentes

El lavado de activos es un delito cometido por una persona natural o jurídica cuando de forma directa o indirecta busca dar apariencia lícita a activos obtenidos de manera ilícita (Junta de Política y Regulación Monetaria y Financiera, 2014). Es considerado como un delito autónomo, y debido a su naturaleza, genera distorsiones en la economía tales como imperfecciones en los mercados financieros y de bienes y servicios. De manera indirecta, genera también imperfecciones en el mercado laboral. En conjunto, todos estos efectos provocan desequilibrios en materia fiscal y cambiaria, distorsionando los resultados de la economía real con repercusiones para los ciudadanos (Unidad de Información y Análisis Financiero, 2014).

El lavado de activos implica la existencia de operaciones anómalas e injustificadas, que en el mercado financiero se evidencian a través de movimientos cuya frecuencia, monto o destinatario no corresponde con el perfil económico y de comportamiento de la persona (Junta de Política y Regulación Monetaria y Financiera, 2014). El alto desarrollo de las telecomunicaciones, acompañado de la interdependencia en materia legislativa de los países, ha permitido la maximización de este fenómeno, cuyos capitales ilícitos fluyen fácilmente en el sistema financiero mundial (Montes, 2014).

De acuerdo con la Oficina de Crimen y Drogas de las Naciones Unidas (UNODC por sus siglas en inglés), se estima que en un año el monto aproximado en lavado de activos representa entre el 2% y 5% del Producto Interno Bruto Global, es decir, entre 800 billones y 2 trillones de dólares americanos. Ésta es una cifra preocupante, aún con la estimación porcentual más baja (United Nations Office on Drugs and Crime, 2019).

En Ecuador, durante el año 2020, la Unidad de Análisis Financiero (UAFE) generó 958 informes ejecutivos como respuesta a requerimientos de información solicitados por la Fiscalía General del Estado (FGE). Del total de dichos informes, el 11% (105) correspondía a delitos de lavado de activos. En el mismo periodo fiscal, la Dirección de Análisis de Operaciones de la UAFE recibió un total de 2.245 reportes de operaciones inusuales e injustificadas (ROII's) por parte de los sujetos obligados a informar, de los cuáles, más del 75% (1.686) fueron reportados por las entidades del sector financiero, las empresas de transferencia de fondos y las cooperativas de ahorro y crédito. Por otro lado, con base a la información de varias carteras del Estado ecuatoriano, la UAFE elaboró un total de 167 informes de operaciones sospechosas (IOS), los cuáles en una etapa posterior podrían convertirse en reportes de operaciones inusuales e injustificadas (Unidad de Análisis Financiero, 2021).

Finalmente, es importante destacar que Ecuador no cuenta con una estimación oficial respecto a la magnitud de los flujos generados por el lavado de activos, no obstante, se estima que el blanqueo de capitales a través de cuentas corrientes es del 4,65% del Producto Interno Bruto real, es decir, más de 4 mil millones de dólares (Bernal & Sares, 2019).

Planteamiento del problema

A nivel mundial, una de las preocupaciones más críticas en materia económica, es la identificación temprana del cometimiento de actividades de lavado de activos en el sistema financiero (Organización de Estados Americanos, 2018). Es por tal motivo que los estados, los organismos de control y las entidades financieras a nivel mundial, buscan constantemente identificar de manera temprana el cometimiento de actividades de lavado de activos en el sistema financiero; y, debido a la gran cantidad de datos que pueden ser analizados, para la generación de alertas tempranas, recurren a alternativas basadas en el uso de tecnologías de la información y comunicaciones (Kashyap, 2018).

En Ecuador, el sistema financiero está compuesto por las instituciones financieras del sector público, privado y popular y solidario (Asamblea Nacional Constituyente, 2008). De acuerdo con la Ley Orgánica de Economía Popular y Solidaria (LOEPS), el control de la Economía Popular y Solidaria (EPS) y del Sector Financiero Popular y Solidario (SFPS) está a cargo de la Superintendencia de Economía Popular y Solidaria (Asamblea Nacional Constituyente, 2011). En esa misma línea, para procesos de control y análisis, las organizaciones bajo la supervisión de la Superintendencia de Economía Popular y Solidaria, reportan periódicamente información relacionada con la situación económica y de gestión; parte de esa información corresponde a los depósitos captados por las entidades del SFPS (Asamblea Nacional Constituyente, 2011).

La información de depósitos es reportada por las entidades del SFPS a la Superintendencia a través de la estructura de depósitos (D01). Durante el año 2020, la Superintendencia de Economía Popular y Solidaria acopió 2.729 estructuras de depósitos, con el detalle de los saldos de los depósitos de los socios y clientes de las entidades del SFPS. En diciembre de 2020, el SFPS tenía 522 entidades, cuyo saldo de depósitos ascendía a USD 14 057 millones (Superintendencia de Economía Popular y Solidaria, 2020).

Dado este escenario, implica que el Sector Financiero Popular y Solidario, al ser parte del sistema financiero nacional, también se encuentra expuesto al fenómeno de lavado de activos. Por ello, dado el ámbito de competencia de la Superintendencia de Economía Popular y Solidaria, debe encaminar acciones que le permitan identificar el mal uso del SFPS para el cometimiento de lavado de activos. Entre las acciones que el organismo de control debe contemplar está la identificación de anomalías en los saldos de los depósitos de las entidades que se encuentran bajo su control (Unidad de Información y Análisis Financiero, 2014). No obstante, dada la cantidad de entidades que se encuentran bajo supervisión del organismo de

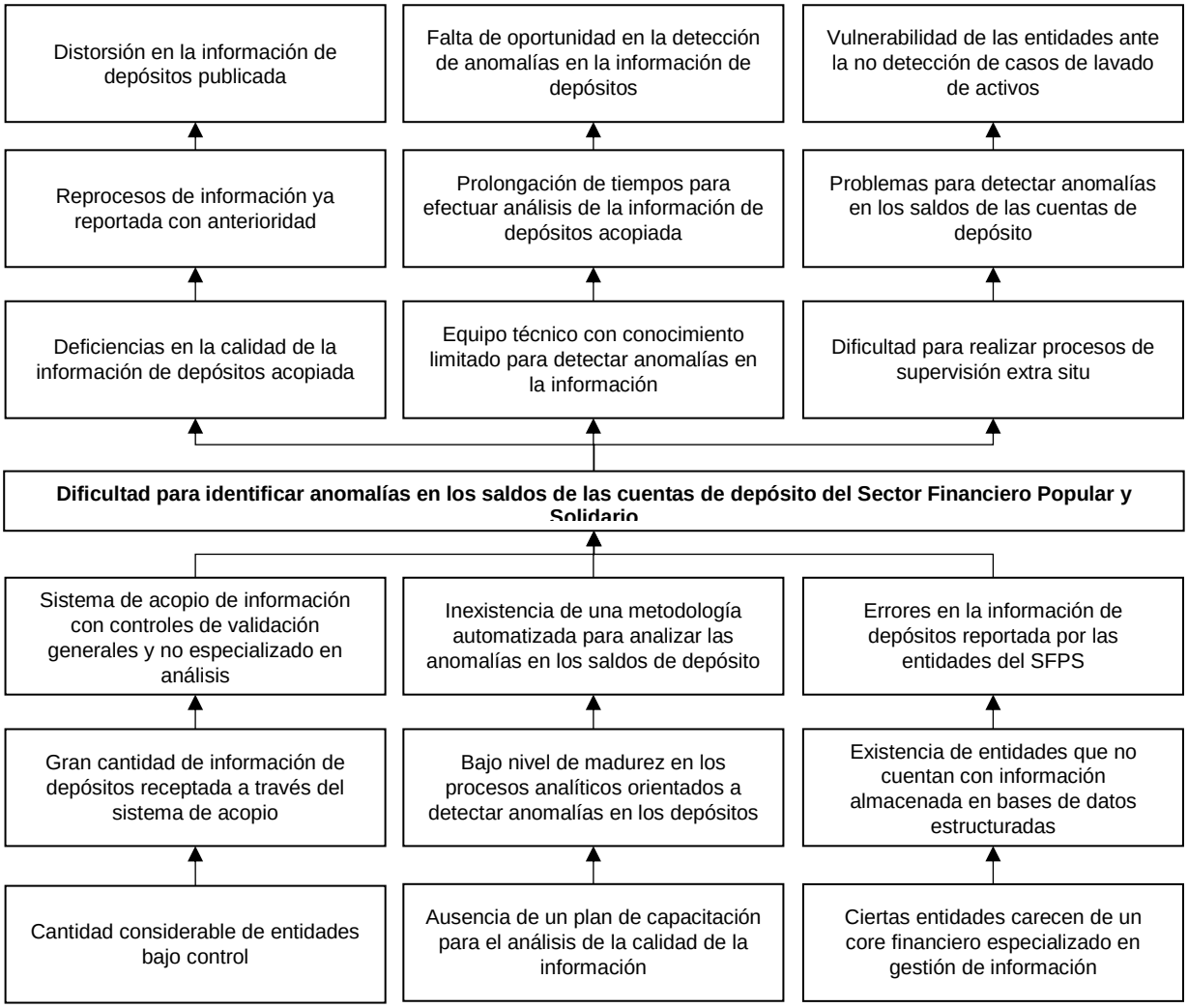
control y a las diferentes características y capacidades de cada una de ellas, la tarea de identificar anomalías en la información de depósitos reportada se torna compleja.

Con esta premisa, teniendo en cuenta que una de las etapas del lavado de activos constituye la colocación de las ganancias ilícitas en el sistema financiero, principalmente a través de cuentas de depósito, surge la siguiente interrogante: ¿podría el organismo de control del Sector Financiero Popular y Solidario identificar anomalías en los saldos de depósitos de las entidades bajo su supervisión, basándose en modelos y técnicas de análisis de datos?

Con estos antecedentes, en la Figura 1, se presentan las causas, el problema central y los efectos de la temática abordada en este trabajo.

Figura 1

Árbol de problemas respecto a la detección de depósitos inusuales en el SFPS



Nota. La figura es una representación gráfica de las causas, problema central y efectos de la temática abordada en este trabajo.

Justificación, importancia y alcance

De acuerdo con el índice de vulnerabilidad ante lavado de activos publicado por el Instituto de Gobernanza de Basilea, al 2019 el Ecuador se ubica en el puesto 29 de un total de

125 países analizados, es decir, entre los países más proclives a verse afectados por prácticas de lavado de activos (Basel Institute on Governance, 2019).

En el caso particular de Ecuador, el Sector Financiero Popular y Solidario representa aproximadamente un tercio de las captaciones del sistema financiero nacional. Por otro lado, el SFPS es bastante heterogéneo en donde las cooperativas del segmento 1 y las asociaciones mutualistas concentran más del 80% del saldo de los depósitos de todo el sector. Los depósitos del SFPS son principalmente depósitos a plazo (69%) llegando a representar aproximadamente el 26% de los depósitos del sistema financiero nacional. A diciembre 2020, de un total de 522 entidades activas, 39 pertenecían al segmento 1 (Superintendencia de Economía Popular y Solidaria, 2020).

El lavado de activos es un fenómeno que comprende varias aristas y actores de la economía; por lo tanto, su mitigación es factible bajo la combinación de múltiples acciones de prevención y detección. En este contexto, la Superintendencia de Economía Popular y Solidaria, como ente regulador del SFPS, entre otras acciones, debe potenciar el uso de la información acopiada a fin de detectar el posible cometimiento de actividades relacionadas al lavado de activos.

Por lo descrito previamente, este trabajo se centra en proponer un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos, como una herramienta de apoyo al Organismo de Control para su análisis ante potenciales casos de lavado de activos, entidad que entre sus principales atribuciones constan el fortalecimiento y la estabilidad del SFPS, a través de sus procesos de acopio de información, monitoreo y supervisión.

Para ello se realizará un análisis de la información de depósitos reportada por las entidades del SFPS al Organismo de Control, y se generará una propuesta de modelo

aplicando técnicas de análisis de datos, cuyos resultados se visualizarán a través de una herramienta de inteligencia de negocios.

Objetivos del proyecto

Objetivo general

Proponer un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del Sector Financiero Popular y Solidario mediante la aplicación de técnicas de análisis de datos a fin de brindar una herramienta de apoyo al análisis de posibles casos de lavado de activos.

Objetivos específicos

OE1. Determinar las causas que dificultan la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del SFPS.

OE2. Realizar una revisión inicial de la literatura para determinar el uso de técnicas de análisis de datos basados en el uso de Tecnologías de información y comunicaciones (TICs) en la identificación anomalías en los saldos de las cuentas de depósitos.

OE3. Construir un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos mediante técnicas de análisis de datos basadas en el uso de TICs como apoyo al análisis de posibles casos de lavado de activos.

OE4. Evaluar la usabilidad del modelo para la identificación de anomalías en los saldos de las cuentas de depósitos como herramienta de apoyo al análisis de posibles casos de lavado de activos.

Preguntas de investigación

Para la consecución de los objetivos previamente descritos, a continuación, se detalla las preguntas de investigación que delimitan el alcance de este trabajo y el estudio del estado del arte a realizar:

OE1.RQ1. ¿Cuáles son las causas que dificultan la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las entidades del SFPS al Organismo de Control y cuáles son sus efectos?

OE1.RQ2. ¿Qué causas y efectos relacionados a la dificultad para identificar anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del SFPS al Organismo de Control serán abordados en este estudio?

OE2.RQ1. ¿Qué tipos de estudios relacionados a la identificación de anomalías en los saldos de las cuentas de depósitos existen?

OE2.RQ2. ¿Cuáles son las técnicas de análisis de datos usadas en la actualidad para identificar anomalías en los saldos de las cuentas de depósitos?

OE2.RQ3. ¿Qué importancia tiene el uso de las TIC's en la identificación de anomalías en los saldos de las cuentas de depósitos?

OE2.RQ4. ¿Qué características presentan las técnicas de identificación de anomalías en los saldos de las cuentas de depósitos basadas en el uso de TIC's?

OE3.RQ1. ¿Con qué fuentes de información se cuenta para realizar el análisis de anomalías en los saldos de las cuentas de depósitos?

OE3.RQ2. ¿Qué técnicas de análisis de datos basados en el uso TIC's serán empleadas para construir el modelo de identificación de anomalías en los saldos de las cuentas de depósitos?

OE3.RQ3. ¿Qué tipos de datos requieren las técnicas de análisis de datos basadas en el uso de TIC's seleccionadas para construir el modelo?

OE3.RQ4. ¿Qué características de visualización para presentar los resultados del modelo son las más adecuadas para el usuario de supervisión y monitoreo?

OE4.RQ1. ¿Qué tan comprensibles son los resultados generados por el modelo para la identificación de anomalías en los saldos de las cuentas de depósitos?

OE4.RQ2. ¿Qué mejoras a futuro al modelo se pueden implementar en la propuesta del modelo de identificación de anomalías en los saldos de las cuentas de depósitos?

Hipótesis

La generación de un modelo mediante la aplicación de técnicas de análisis de datos basadas en el uso de TIC's permitirá identificar anomalías en los saldos de las cuentas de depósito en la información que es reportada por las entidades del Sector Financiero Popular y Solidario (SFPS) al organismo de control.

Capítulo II

Marco teórico

Antecedentes

Como se destacó en el Capítulo I, el lavado de dinero es un fenómeno de índole mundial y, por tanto, una preocupación de todo Estado. Además, se enfatizó que uno de los principales sectores económicos afectados por esta problemática es el sector financiero, debido a su rol de intermediación financiera. Dado este escenario, son varios los estudios realizados a nivel nacional e internacional respecto a un potencial uso de técnicas de análisis de datos para la detección de fraude, lavado de dinero y delitos relacionados, sobre todo considerando que, a la par del avance tecnológico, también han aumentado de forma exponencial la cantidad de datos que se requieren analizar. En este sentido, a continuación, se procede a citar las conclusiones de algunos de estos trabajos de investigación.

Bouazza, Ameur, & Ameur (2019) realizaron un estudio del arte respecto a la detección de fraude financiero a través del uso de técnicas de minería de datos. El periodo analizado comprendió desde el año 1966 hasta el año 2017. Entre las conclusiones, los autores destacaron la aplicabilidad de las técnicas de minería de datos en diferentes dominios, y en el caso puntual del fraude financiero como una herramienta que puede ayudar a detectar y anticipar actos de fraude de tal manera que se pueda adoptar medidas para minimizar su impacto. Además, recalcaron el uso de modelos logísticos, redes neuronales, redes bayesianas y árboles de decisión, como técnicas aplicables para detectar y clasificar datos relacionados al cometimiento de fraudes.

En su artículo denominado ‘Potenciales aplicaciones de la minería de datos en el Ecuador’, Camana (2016) concluyó que, la minería de datos es una herramienta de apoyo que permite disminuir la brecha existente entre la cantidad de datos y la capacidad humana para el

procesamiento de dicha información, facilitando los procesos de exploración, análisis, conocimiento y aplicación del conocimiento obtenido de los grandes volúmenes de datos.

En el estudio denominado ‘Metodologías de aprendizaje automático contra el blanqueo de capitales en corresponsales no bancarios’, Guevara, García, & Granados (2020) concluyeron como relevante el uso de algoritmos de visualización y aprendizaje automático para la identificación de anomalías en las transacciones de corresponsales no bancarios, principalmente en países en vías de desarrollo tales como Colombia, México y Ecuador. Además, plantearon como temáticas para estudios futuros los siguientes: i) análisis focalizado en ciudades con gran presencia de corresponsales no bancarios, ii) confrontación de resultados entre técnicas de aprendizaje supervisado y no supervisado, iii) uso de técnicas de modelado gráfico probabilístico (PMG por sus siglas en inglés); y, iv) profundización de análisis de anomalías a través de la segmentación de clústeres.

En el análisis de los procedimientos utilizados por la auditoría forense aplicada a la prevención de lavado de activos en el sector de la banca privada en la ciudad de Cuenca, Pesántez (2020) concluyó que la aplicación de la inteligencia artificial y las técnicas de minería de datos son elementos fundamentales en el proceso de identificación de transacciones inusuales en las instituciones financieras. Además, destacó que la inteligencia artificial y la minería de datos facilitan la identificación de señales de alerta, identificación de casos atípicos y toma de decisiones.

Mediante la implementación de un modelo analítico para la identificación de comportamientos inusuales en los pagos anticipados de una entidad de arrendamiento financiero, Hernández (2019) determinó que las herramientas analíticas, las técnicas estadísticas y los modelos computacionales, al permitir el procesamiento de grandes volúmenes de información, facilitan la detección de problemáticas relacionadas al lavado de activos y la financiación del terrorismo. En la misma línea, destacó el uso de técnicas

estadísticas de clasificación para la caracterización de valores atípicos e identificación de comportamientos inusuales.

Mejía (2018) construyó un modelo para la identificación de patrones que influyen en la mora de una cooperativa de ahorro y crédito del Ecuador, a través del uso de técnicas de minería de datos. En su trabajo destacó el uso de algoritmos tales como las redes neuronales, clustering y árboles de decisión.

Finalmente, Rojas (2018) en su trabajo de titulado ‘Propuesta metodológica para la detección y prevención de fraudes de lavado de activos en empresas del sector inmobiliario empleando herramientas de análisis de datos lógicos’, sacó a relieve la importancia del uso de herramientas de análisis de datos en la ejecución de procesos de auditoría, pues le permiten al auditor optimizar recursos y enfocarse con claridad en posibles esquemas de lavado de activos.

En complemento a los estudios previamente citados, este trabajo tiene como finalidad, en primera instancia, la construcción de un modelo basado en el uso de técnicas de minería de datos que permita detectar anomalías en los saldos de los depósitos de las entidades que conforman el Sector Financiero Popular y Solidario. Además, busca validar la usabilidad de las técnicas de minería de datos, considerando las características de la información y la heterogeneidad (en cuanto al tamaño) de las entidades que se encuentran bajo el control y la supervisión de la Superintendencia de Economía Popular y Solidaria.

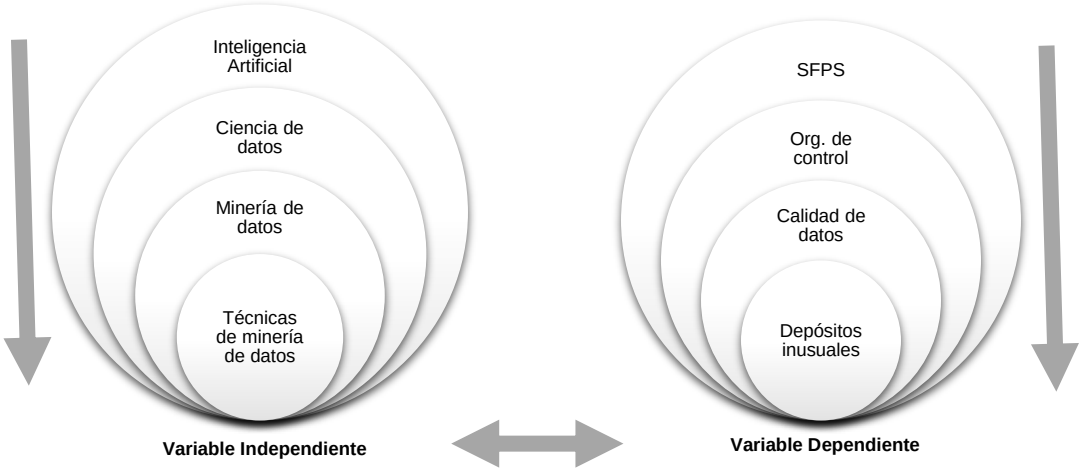
Con estos antecedentes, a continuación, se procede a exponer la red de categorías y la fundamentación teórica de las variables dependiente e independiente abordadas en este trabajo.

Red de categorías

Con el objetivo de buscar congruencia entre los conceptos teóricos y la hipótesis previamente planteada, en la Figura 2, se presenta de forma jerárquica a la red de categorías de las variables.

Figura 2

Red de categoría de variables



Nota. La figura es una representación gráfica de la red de categorías de las variables independiente y dependiente que enfocan el curso de este trabajo. SFPS son las siglas de Sector Financiero Popular y Solidario.

Fundamentación de la variable independiente

La **Inteligencia Artificial** es una ciencia multidisciplinaria cuyo origen se remonta a más de 60 años, en la que se interactúan áreas del conocimiento como la matemática, informática, psicología y biología; y cuya principal finalidad es proporcionar a los sistemas computacionales de características que simulen la inteligencia. En la actualidad, el empoderamiento de la Internet de las Cosas (IoT) y la generación de grandes volúmenes de datos (Big Data) han impulsado aún más el desarrollo de la Inteligencia Artificial. A nivel mundial, esta ciencia está presentando mayor importancia en áreas como la salud, finanzas, gobierno y educación.

Debido a su amplio espectro de aplicación, la inteligencia artificial se desagrega en varias ramas de conocimiento como: aprendizaje automático, ciencia de datos, procesamiento de lenguaje natural, sistemas expertos, visión, lenguaje, planificación y robótica (Haugeland, 1988).

La ciencia de datos, comúnmente conocida como **Data Science**, es una combinación de técnicas que extraen valor de los datos. Algunas de las técnicas usadas en la ciencia de datos tienen su origen en principios de estadística aplicada, aprendizaje automático, visualización, lógica y ciencias de la computación. La ciencia de datos es principalmente utilizada con fines predictivos y para el descubrimiento de patrones y relaciones, y aplicada en múltiples disciplinas como finanzas, marketing, ingeniería, entre otras. Hoy en día, la ciencia de datos se ha convertido en una herramienta esencial para toda organización que captura, almacena y procesa datos como parte de sus operaciones (Kotu & Deshpande, 2019).

Dentro de la ciencia de datos, el proceso a través del cual se descubren relaciones, patrones y tendencias basadas en el análisis de grandes cantidades de datos, se denomina **Minería de Datos** (Data Mining). La existencia de grandes volúmenes de datos, acompañado del uso de tecnologías de la información y comunicación, ha transformado el análisis de datos orientándolo hacia la aplicación de determinadas técnicas especializadas (Pérez, 2007).

Las **técnicas de minería de datos** se sustentan primordialmente en la inteligencia artificial y la estadística, que en conjunto implican el desarrollo de algoritmos más o menos sofisticados que se emplean sobre un conjunto de datos para obtener como resultado información y conocimiento. En términos generales, las técnicas de minería de datos son de tipo descriptivas y predictivas. La elección de la técnica de minería de datos depende de la tarea que se va a realizar en el caso de estudio. Las técnicas predictivas definen el modelo para el análisis de los datos en base a un conocimiento teórico previo, el cual posteriormente debe contrastarse para ser considerado como válido. En contraparte, las técnicas descriptivas

no asignan ningún papel predeterminado a las variables, es decir, no supone la existencia de variables dependientes e independientes, ni tampoco la existencia de un modelo previo para el análisis de los datos, por lo cual los modelos se crean automáticamente partiendo del reconocimiento de patrones en los datos que son objeto de análisis (Pérez, 2007).

Dentro de las técnicas de minería de datos predictivas están: regresiones, series temporales, análisis de varianza y covarianza, análisis discriminante, árboles de decisión, redes neuronales, algoritmos genéticos; y, técnicas bayesianas. Finalmente, en el grupo de las técnicas de minería de datos descriptivas tenemos: clustering y segmentación, técnicas de asociación y dependencia, técnicas de análisis exploratorio de datos; y, técnicas de reducción de dimensión y escalamiento multidimensional (Vieria, 2009).

Fundamentación de la variable dependiente

La Constitución de la República del Ecuador, en su artículo 283, determina que el sistema económico de Ecuador es social y solidario y está integrado por las formas de organización económica de tipo pública, privada, mixta, popular y solidaria, y las demás que la Constitución establezca. Además, define que el sector de la economía popular y solidaria estará regulada según la normativa establecida para el efecto y estará conformado por el sector cooperativista, asociativo y comunitario. En el artículo 309 del referido cuerpo legal, se determina que el **sistema financiero nacional** está compuesto por el sector público, privado y **popular y solidario**, que contarán con normas y entidades de control específicas para garantizar la seguridad, estabilidad, transparencia y solidez de dichos sectores (Asamblea Nacional Constituyente, 2008).

La Ley Orgánica de Economía Popular y Solidaria (LOEPS), en su artículo 146, señala que la **Superintendencia de Economía Popular y Solidaria** (SEPS) es un organismo técnico, con jurisdicción nacional y con personalidad jurídica de derecho público, quien tendrá la potestad de ejercer el control de la Economía Popular y Solidaria (EPS) y del Sector Financiero

Popular y Solidario (SFPS). Entre las principales atribuciones de la SEPS están: i) ejercer la supervisión de las actividades económicas de las personas y organizaciones sujetas a la LOEPS, ii) velar por la estabilidad, solidez y correcto funcionamiento de las instituciones sujetas a su control, iii) levantar estadísticas de las actividades que realizan las organizaciones sujetas a la Ley, entre otras (Asamblea Nacional Constituyente, 2011).

La Superintendencia de Economía Popular y Solidaria, de acuerdo con lo establecido en su Estatuto Orgánico de Gestión Organizacional por Procesos Institucional, está conformada por procesos gobernantes, procesos adjetivos y procesos sustantivos (Superintendencia de Economía Popular y Solidaria, 2019). (Ver Figura 3)

Figura 3

Mapa de procesos de la SEPS

Nota. La figura es un esquema resumen del mapa de procesos de la SEPS. Elaborado a partir del Estatuto Orgánico de Gestión Organizacional por Procesos, Superintendencia de Economía Popular y Solidaria, 2019.

Dentro de los procesos sustantivos se encuentra la Gestión General de Servicios e Inteligencia de la Información, que a su vez está compuesta por la Gestión de Servicios de la

Economía Popular y Solidaria; y, la Gestión de Información y Normativa Técnica. La misión del proceso de Gestión de Información y Normativa Técnica es, entre otros, orientar los procesos de gestión y calidad de la información, así como la producción de estadísticas y estudios, relacionados con la Economía Popular y Solidaria, contribuyendo al desarrollo y estabilidad de los sujetos bajo control. Además, figuran como atribuciones y responsabilidades del proceso de Gestión de Información y Normativa Técnica, el monitoreo de los estándares de calidad de la información reportada por las entidades del sector financiero popular y solidario y las organizaciones de la economía popular y solidaria (Superintendencia de Economía Popular y Solidaria, 2019).

La **calidad de los datos** es una de las principales preocupaciones de las organizaciones que trabajan con grandes volúmenes de datos, ya que ésta es a la vez el principal activo con el que cuentan para apoyar su proceso de toma de decisiones (Echegoyen, 2003). La calidad es la cualidad de la información recogida a través de una base de datos o sistema de información, que presenta atributos como:

- Pertinencia: que se refiere al grado en el cual los datos satisfacen las necesidades de los usuarios actuales y potenciales.
- Precisión y confiabilidad: que es la medida en la que los datos están libres de errores causados por diversos factores.
- Oportunidad y puntualidad: se refiere a la velocidad con que se publican los datos respecto a una calendarización definida.
- Accesibilidad y claridad: es la facilidad con la que los usuarios pueden acceder a los datos y la medida en que se explican a través de los metadatos.
- Comparabilidad: se refiere a la medida en que los datos pueden compararse a través del tiempo u otra dimensión categórica.

- Coherencia: se refiere al grado en el que los datos se ajustan a las definiciones, controles y metodologías establecidas.

La calidad de los datos se vuelve muy relevante en procesos de supervisión, para los cuáles la SEPS, a través de su sistema de acopio recepta varias estructuras de información, entre ellas: estados financieros, depósitos, cartera, inversiones, socios y servicios financieros. En este punto es relevante puntualizar que el presente trabajo busca centrarse en la detección de anomalías en la calidad y consistencia de información de los saldos de las cuentas depósito.

Desde una perspectiva de registros, la existencia de **anomalías en los saldos de las cuentas de depósitos**, es decir, los movimientos de dinero realizados por personas naturales o jurídicas, cuya frecuencia, monto o destinatario no guarda correspondencia con su perfil económico y de comportamiento; pueden significar potenciales casos inmersos en lavado de activos (Junta de Política y Regulación Monetaria y Financiera, 2014). El lavado de activos se puede identificar en tres etapas:

- Colocación: es el proceso a través del cual se introducen las ganancias ilícitas en el sistema financiero.
- Estratificación: en esta etapa los fondos se separan en múltiples transacciones con el objetivo de evitar su rastreo.
- Integración: los fondos ilegales son reinsertados en la economía, dando la apariencia de ser lícitos.

Bajo este contexto, se pueden catalogar como operaciones de depósito anómalas a:

- Alteración al nivel de transacciones en las cuentas.
- Incremento de la cantidad de dinero.
- Transacciones con movimientos significativos.

Estado del arte

El análisis del estado del arte relacionado con la identificación de depósitos anómalos en el marco de casos de lavado de activos se ha realizado mediante la aplicación de una revisión inicial de la literatura, en la cual se ha seguido los siguientes pasos: la determinación de los criterios de inclusión y exclusión, estructuración del grupo de control, definición de la cadena de búsqueda ideal, selección de estudios y síntesis de los resultados obtenidos. Todo este proceso se lo realizó en la base digital Springer.

Determinación de los criterios de inclusión y exclusión

Los criterios de inclusión y exclusión permiten ubicar investigaciones realizadas por otros autores, en función a las características determinadas para la el análisis. Para este proceso, se definen los siguientes:

Criterios de inclusión

- Artículos cuyo propósito sea la aplicación de técnicas de análisis de datos basados en el uso de TICs para la detección de anomalías en los saldos de las cuentas de depósitos.
- Artículos que evidencien la validez de la aplicación de técnicas de análisis de datos basadas en el uso de TICs como una herramienta adicional para el análisis del cometimiento de actos ilícitos a través de los depósitos en el sistema financiero.
- Artículos que propongan una metodología de aplicación de técnicas de análisis de datos basadas en el uso de TICs para el control de la calidad y la consistencia de la información.

Criterios de exclusión

- Artículos enfocados exclusivamente en el proceso de implementación tecnológico de modelos de análisis de datos.
- Artículos enfocados en la disminución del cometimiento de actos ilícitos de lavado de dinero más no en la forma de aprovechar las tecnologías de la información para su detección y control.

Estructura del grupo de control

De acuerdo con una revisión previa, se han seleccionado 3 artículos técnicos de la base de datos digital Springer, cuyas palabras clave se presentan en la Tabla 1:

Tabla 1

Términos clave

#	Referencias	Términos
S1	Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review (Chen, y otros, 2018).	Anti-money laundering, Machine learning, Data mining, Algorithms, Supervised learning, Unsupervised learning, Anomaly detection, Behavioral modelling, Risk scoring
S2	Data Mining Techniques Applied in the Financial Industry (Huo, Wang, & Liu, 2013)	Data mining, Financial data, Prediction, Discover techniques
S3	Detecting Frauds and Money Laundering: A Tutorial (Keshav Palshikar, 2014)	Money laundering, Fraud detection, Analytics techniques.

Definición de la cadena de búsqueda ideal

De los estudios técnicos del grupo de control revisado en la sección anterior, se determinaron 16 términos clave, una vez eliminados los duplicados. Con ello, se realiza un pilotaje en la base de datos digital Springer a fin de determinar la cadena de búsqueda ideal. Los resultados obtenidos del proceso de búsqueda en la base de datos digital se pueden visualizar en la Tabla 2.

Tabla 2

Cadena de búsqueda

#	Cadena	Resultado	Grupo de control
1	("MONEY LAUNDERING") AND ("DEPOSIT" OR "SAVING" OR "ACCOUNT") AND "FINANCIAL SYSTEM"	932	S1, S2 y S3
2	"MONEY LAUNDERING" AND ("DEPOSIT" OR "SAVING" OR "ACCOUNT") AND ("SUSPICIOUS" OR "ANOMALY") AND "DETECTION" AND "FINANCIAL SYSTEM"	150	S1, S2 y S3
3	"MONEY LAUNDERING" AND ("SUSPICIOUS" OR "ANOMALY") AND ("DEPOSIT" OR "SAVING" OR "ACCOUNT") AND ("ACTIVITY" OR "OPERATION" OR "PATTERN" OR "TRANSACTION") AND "DETECTION" AND "FINANCIAL SYSTEM"	150	S1, S2 y S3

#	Cadena	Resultado	Grupo de control
4	"MONEY LAUNDERING" AND ("SUSPICIOUS" OR "ANOMALY") AND ("DEPOSIT" OR "SAVING" OR "ACCOUNT") AND ("ACTIVITY" OR "OPERATION" OR "PATTERN" OR "TRANSACTION") AND ("DETECT" OR "IDENTIFY") AND ("DATA ANALYSIS" OR "ANALYTICS") AND "FINANCIAL SYSTEM"	53	S1, S2 y S3
5	"MONEY LAUNDERING" AND ("SUSPICIOUS" OR "ANOMALY") AND ("DEPOSIT" OR "SAVING" OR "ACCOUNT") AND ("ACTIVITY" OR "OPERATION" OR "PATTERN" OR "TRANSACTION") AND ("DETECT" OR "IDENTIFY" OR "FIND") AND ("DATA ANALYSIS" OR "ANALYTICS") AND ("TECHNIQUE" OR "MODEL") AND ("FINANCIAL SYSTEM" OR "BANK" OR "COOPERATIVE")	26	S1

De acuerdo con la Tabla 2, la cadena ideal es la número cuatro puesto que permite obtener la menor cantidad de artículos, incluyendo a todos los estudios del grupo de control.

Selección de estudios

A partir de los resultados obtenidos en el paso previo y los criterios de inclusión y exclusión determinados de forma preliminar, a continuación, en la Tabla 3 se detallan los estudios primarios seleccionados.

Tabla 3

Estudios seleccionados

#	Nombre del estudio
1	Data Science Applications
2	Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review
3	Detecting Frauds and Money Laundering: A Tutorial Integrating Client Profiling in an Anti-money Laundering Multi-agent Based System
4	Sequence Matching for Suspicious Activity Detection in Anti-Money Laundering
5	Technology Stack for Machine Learning and Associated Technologies
6	Data Mining Techniques Applied in the Financial Industry

Síntesis de resultados obtenidos

Los 6 estudios seleccionados, están comprendidos entre el año 2008 y el año 2018. A continuación, se exponen las ideas principales obtenidos de ellos:

- En principio, la ciencia de datos puede ser aplicada a cualquier campo de estudio o área, es así como, en la actualidad, es empleada para el análisis de problemáticas y generación de soluciones en el sistema financiero y la sociedad. Las instituciones financieras internacionales han implementado soluciones tales como gestión de riesgo, prevención de actividades fraudulentas, lavado de activos, en donde las técnicas de minería de datos son

consideradas como adecuadas para la correcta identificación de este tipo de actividades (Cao, Longbing, 2018).

- Grandes cantidades de dinero se lavan cada año, representando no solo una amenaza para la estabilidad de las instituciones financieras sino también para la seguridad de la economía mundial. En este marco, el machine learning se presenta como una herramienta exhaustiva de aplicación de algoritmos de aprendizaje automático para la identificación de transacciones sospechosas identificando soluciones de tipología contra el lavado de dinero tales como: modelos de comportamiento, análisis de relaciones, calificación de riesgo, detección de anomalías y capacidad geográfica. Se presenta además una secuencia de pasos lógica para la preparación, transformación y el análisis de los datos (Chen, y otros, 2018).
- Las instituciones financieras pueden obtener mejores resultados en las iniciativas antilavado de activos, mediante la mejora el proceso de señalización de transacciones sospechosas y la decisión final posterior. Para esto, es crítico modelar el comportamiento del cliente, teniendo una definición clara de los diferentes perfiles del cliente. En este sentido, mediante un conjunto de datos bancarios del mundo real, se explica en esta contribución cómo se utilizaron algunas técnicas de extracción de datos para crear los perfiles de clientes necesarios y cómo los resultados obtenidos pueden integrarse en un sistema de alertas tempranas. Los cinco algoritmos principales capaces de modelar eventos de lavado de activos son: la regresión logística de Bayes, los árboles de decisión, los bosques aleatorios, las máquinas de soporte vectorial y las redes neuronales artificiales. En adición, debido a que los eventos de lavado de dinero son raros, a fin de detectar dichas anomalías, los algoritmos deben combinarse con técnicas de muestreo y clasificación de eventos (Keshav, 2014).
- El desarrollo de métodos efectivos para detectar operaciones de lavado de activos se ha convertido en la actualidad en una prioridad para las instituciones financieras y el gobierno.

Si bien existen múltiples algoritmos que permiten detectar anomalías, un modelo de alta precisión puede obtenerse la selección de secuencias de referencia de dos tipos: 1) el historial de transacciones de la cuenta individual y 2) la información de transacciones de otras cuentas en un grupo de pares. Esta metodología busca detectar anomalías de comportamientos aparentemente normales (Xuan, Pengzhu, & Zeng, 2008).

- El vertiginoso desarrollo de las Tecnologías de Información también significó la ampliación del espectro de actuación de actividades fraudulentas, en consecuencia, el software que en su momento fue diseñado para detectar operaciones anómalas, debe ser rediseñado mediante la implementación de algoritmos de minería de datos para mejorar su precisión (Kashyap, 2018).
- La detección de fraudes bancarios en línea incorpora varias técnicas avanzadas de extracción de datos, por ejemplo, al crear un vector de contraste para cada transacción en función de la secuencia de comportamiento histórico de un individuo, se puede perfilar la tasa de diferenciación de cada transacción actual contra la preferencia de comportamiento del mismo individuo. Para este fin, se puede utilizar un algoritmo denominado ContrastMiner, que extrae los patrones de contraste y distingue el comportamiento fraudulento del genuino, seguido de una selección de patrones efectiva y un puntaje de riesgo que combina predicciones de diferentes modelos (Wei, Jin, Longbing, & Chen, 2012).

En base a lo anterior, se puede concluir que, el arte y la ciencia de detectar intrusiones y fraudes en el monitoreo y las transacciones de datos se remonta a más de 25 años atrás; y, en el ámbito financiero, las diferentes técnicas englobadas dentro de lo que se conoce como ciencia de datos, tiene múltiples aplicaciones. Los enfoques para la detección de anomalías pueden clasificarse según el tipo de comportamiento modelado y dicho enfoque asume: i) Que el comportamiento normal, ya sea el comportamiento de un usuario, un protocolo o proceso, puede ser descrito con precisión, por ejemplo, estadísticamente; ii) Que las anomalías se

desvíen del comportamiento normal, y iii) Que todas las desviaciones del comportamiento normal representan anomalías. Las herramientas para la detección de comportamientos anómalos se catalogan en: genéricas, especializadas y personalizadas, con un amplio uso de técnicas de minería de datos.

Capítulo III

Metodología de la investigación

Enfoque de la investigación

La investigación es un proceso que tiene como finalidad conseguir información que permita validar o corregir el conocimiento existente de un área de estudio, para ello, la investigación es llevada a cabo a través de la aplicación de métodos de tipo científico o experimental. Para cumplir el objetivo mencionado, se requiere contemplar el diseño de una 'metodología de la investigación', la cual es un conjunto de técnicas y procedimientos que se aplican de forma ordenada y sistemática en un estudio. En una investigación, la metodología otorga vigor y severidad científica a los resultados alcanzados en el estudio (Namakforoosh, 2000).

La metodología de investigación puede ser de tipo cuantitativa y cualitativa. La metodología cuantitativa es aquella que se fundamenta sobre datos cuantificables, es decir, obtenidos de procesos como la medición y observación, y que a través de cálculos estadísticos, permiten obtener resultados y conclusiones. La metodología cualitativa es aquella que se fundamenta principalmente sobre la observación directa, entrevistas y análisis; es decir, aplica procedimientos analíticos e interpretativos para abordar el objeto del estudio (Rojas, 2016).

En función de lo descrito en líneas anteriores, el proyecto abordado en este trabajo que se denomina 'Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos en el sector financiero popular y solidario del Ecuador', cuyo objetivo principal es 'Proponer un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del Sector Financiero Popular y Solidario mediante la aplicación de técnicas de análisis de datos a fin de brindar una

herramienta de apoyo al análisis de posibles casos de lavado de activos', se llevará a cabo mediante la aplicación de un enfoque de ***investigación cuantitativo***.

Planteamiento de la metodología de la investigación

El valor de la analítica aplicada a los datos es ampliamente aceptado como relevante por el sector industrial y la comunidad académica, por tanto, es fundamental seguir un proceso claro para abordar problemas relacionados a la gestión de datos (Jagadish, et al., 2014).

Fayyad, Piatetsky-Shapiro & Smyth (1996) definieron al proceso de descubrimiento de conocimiento en las bases de datos (KDD, del inglés Knowledge Discovery in Databases) como un proceso no trivial, entendible, con potencial de uso, novedoso y válidos, a través del cual se identifican patrones en los datos.

El término KDD ha evolucionado a través del tiempo siendo así que en la actualidad la terminología denominada 'Descubrimiento de conocimiento y minería de datos' (KDDM por sus siglas en inglés) es comúnmente empleada para referirse al proceso de descubrimiento de conocimiento aplicado a cualquier fuente de datos. KDDM contempla un proceso completo de extracción de conocimiento que incluye: el almacenamiento y acceso a los datos, la identificación e implementación eficiente de algoritmos que puedan ser usados para analizar grandes cantidades de datos, la interpretación y visualización de los resultados, y la interacción entre el ser humano y la máquina (Kurgan & Musilek, 2006).

Existen varias metodologías que permiten abordar y ejecutar con éxito proyectos de minería de datos, sin embargo, CRISP-DM, SEMMA y KDD son las metodologías que tradicionalmente se emplean para este fin. En la Tabla 4 se presenta un resumen comparativo de las fases que componen cada una de estas tres metodologías generalmente utilizadas para desarrollar proyecto de minería de datos.

Tabla 4

Resumen comparativo de las fases de las metodologías KDD, SEMMA y CRISP-DM

KDD	SEMMA	CRISP-DM
Pre KDD	----------	Comprensión del negocio
Selección	Muestreo	Entendimiento de los datos
Preprocesamiento	Exploración	
Transformación	Modificación	Preparación de los datos
Minería de datos	Modelado	Modelado
Interpretación y evaluación	Evaluación	Evaluación
Post KDD	----------	Implementación

Nota. La tabla presenta un resumen comparativo de las fases correspondientes a las metodologías KDD, SEMMA y CRISP-DM. Tomado de KDD, semma and CRISP-DM: A parallel overview, 2008.

Como se puede apreciar en la Tabla 4, las metodologías KDD, SEMMA y CRISP-DM, independientemente de la diferencia existente en los nombres de las fases que conforman cada metodología, abordan la construcción de una solución basada en minería de datos contemplando etapas como el entendimiento y la preparación de los datos, la minería de los datos y la evaluación e interpretación de los resultados.

Conceptualmente, la metodología CRISP-DM se diferencia de las metodologías KDD y SEMMA, debido a que CRISP-DM contempla dos etapas adicionales relacionadas con la comprensión del negocio e implementación de la solución; sin embargo, esto no significa que el empleo de las metodologías KDD y SEMMA sea limitado, pues toda metodología es susceptible de adaptarse de acuerdo con el proyecto de investigación (Azevedo, 2008).

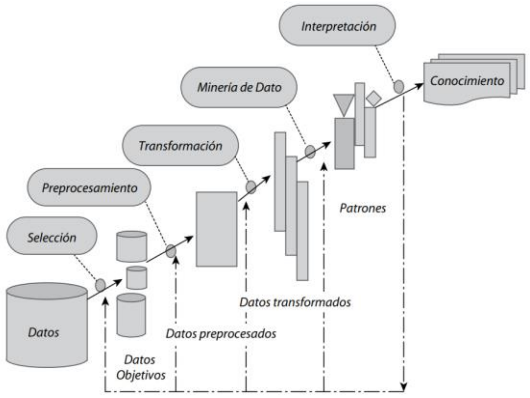
En función con lo descrito en líneas previas y tendiendo presente que el objetivo del presente trabajo es ‘Proponer un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del Sector Financiero Popular y Solidario mediante la aplicación de técnicas de análisis de datos a fin de brindar una herramienta de apoyo al análisis de posibles casos de lavado de activos’, este proyecto será abordado con la metodología de ‘Descubrimiento de Conocimiento en Bases de Datos’ comúnmente conocida como KDD.

Descubrimiento de Conocimiento en Bases de Datos

El descubrimiento de conocimiento en bases de datos es un proceso que busca extraer patrones en forma de reglas o funciones para un análisis posterior por parte del usuario, y abarca tareas como el preprocesamiento de los datos, la minería de datos y la presentación de resultados. KDD es un proceso que permite la iteración e interacción entre sus etapas como se presenta en la Figura 4.

Figura 4

Etapas del proceso KDD



Nota. La figura es una representación gráfica de las etapas del proceso KDD. Tomado de El proceso de descubrimiento de conocimiento en bases de datos, Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, & Hidalgo-Troya, 2016

Etapa de selección

Es la etapa inicial del proceso KDD en la cual se identifica el conjunto de datos a ser empleado para la tarea de descubrimiento. La selección del conjunto de datos puede ser parcial o una muestra representativa de este. Debido a que la selección de los datos varía de acuerdo con los objetivos específicos del negocio, es sumamente relevante que previo a la ejecución de esta etapa, estén claramente definidos los objetivos KDD desde el punto de vista del usuario final (Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, & Hidalgo-Troya, 2016).

Etapa de preprocesamiento

También conocida como etapa de limpieza, consiste en un proceso a través del cual se evalúa la calidad y consistencia de los datos, de tal manera que se obtenga un conjunto de datos adecuado para la aplicación de las tareas de minería de datos (Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, & Hidalgo-Troya, 2016).

De manera general esta etapa implica:

- Identificar datos duplicados,
- Eliminar datos atípicos o ruidosos,
- Emplear estrategias para el manejo de datos perdidos o sin información,

Etapa de transformación

De acuerdo con Fayyad, Piatetsky-Shapiro & Smyth (1996) esta etapa tiene como finalidad buscar características útiles que, en función de la meta KDD, aporten a la representación de los datos. Este proceso se lleva a cabo mediante el empleo de métodos de reducción de dimensiones o de transformación. A continuación, se explica en qué consisten estos dos métodos:

- La **reducción de dimensiones** busca disminuir o simplificar el tamaño del conjunto de datos a ser empleado en la tarea de descubrimiento, y puede ser de tipo horizontal o vertical. Una reducción horizontal es aquella en la que se disminuye el número de registros o filas a través de procesos de sustitución de valores de bajo nivel por agrupaciones de alto nivel, o a través de un proceso de discretización de valores continuos. En cuanto a la reducción vertical, es aquella en donde se suprimen o excluyen de columnas o atributos que: i) no aportan valor para el cumplimiento de la meta KDD o ii) son redundantes. Algunas de las técnicas empleadas para la reducción de dimensiones son agregaciones, segmentaciones, discretización, muestreo, entre otras.
- La **transformación** tiene como finalidad modificar la estructura de los campos o valores de los registros sin que ello implique una alteración de la capacidad para analizar la base de datos objeto de la meta KDD. Generalmente la transformación implica la construcción o discretización de los atributos para mejorar el poder de análisis de los datos o porque se requiere sean ajustados para ser empleados en el algoritmo de minería de datos.

Etapa de minería de datos

Esta etapa tiene como finalidad identificar y descubrir patrones de interés a través de la aplicación de tareas de descubrimiento como la clasificación, el clustering, patrones secuenciales, asociaciones, entre otras (Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, & Hidalgo-Troya, 2016).

Como resultado de la aplicación de las técnicas de minería de datos, se obtienen los modelos, los cuales dependiendo de su objetivo pueden ser predictivos o descriptivos. Los modelos predictivos tienen como principal objetivo la estimación o predicción de un valor futuro o desconocido, a través del empleo de tareas de regresión o clasificación. Por otro lado, los

modelos descriptivos se caracterizan por buscar identificar patrones que expliquen o resuman los datos, apoyándose para ello en el uso de técnicas tales como las reglas de asociación, los patrones secuenciales, el clustering y las correlaciones.

En resumen, esta etapa consiste en i) seleccionar la tarea de minería de datos (descriptiva o predictiva), ii) determinar el o los algoritmos que serán aplicados en la solución y iii) el uso de el o los algoritmos.

Etapa de interpretación

La fase de interpretación-evaluación tiene como finalidad determinar si los patrones descubiertos son precisos, comprensibles e interesantes para el usuario, y suele implicar la iteración con etapas anteriores para modificar el algoritmo, refinar el conjunto de datos o incluso redefinir la meta KDD (Azevedo, 2008).

Por lo general, el conocimiento descubierto es consolidado e incorporado en algún sistema, o simplemente es documentado y reportado a las partes interesadas (Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, & Hidalgo-Troya, 2016).

Dado que con frecuencia los modelos obtenidos no son de fácil lectura para el usuario de negocio, generalmente esta etapa suele incluir la visualización y traducción de los patrones obtenidos en términos entendibles para el usuario final.

Finalmente, con el objetivo de determinar que el conocimiento obtenido es preciso en función de la meta KDD establecida, para la evaluación del modelo suele emplearse un conjunto de datos de entrenamiento (training-set) y un conjunto de datos de prueba (test-set). La evaluación de los resultados del modelo puede ser mediante procesos de:

- Validación simple, que consiste en emplear una parte de la base de datos como conjunto de prueba y el resto como conjunto de entrenamiento.

- Validación cruzada, la cual es empleada cuando se dispone de pocos datos y consiste en dividir la base de datos en dos conjuntos del mismo tamaño. Ambos conjuntos de datos actúan como conjunto de datos de entrenamiento y de prueba de forma simultánea.
- Validación cruzada con n pliegues, que es una variante de la validación cruzada y consiste en dividir la base de datos en n conjuntos de datos. En cada iteración uno de los n conjuntos de datos actúa como conjunto de entrenamiento y el resto como conjuntos de prueba.

Procesamiento de datos y visualización de resultados

Para el procesamiento de las bases de datos utilizadas en este proyecto se emplean las siguientes herramientas:

- R versión 4.1.2, el cual es un software libre diseñado para el procesamiento de información y generación de análisis estadístico.
- R Studio, el cual es un entorno de desarrollo integrado (IDE), que permite manejar R desde un ambiente visual más amigable con el usuario.
- Power BI, que es una herramienta de inteligencia de negocios especializado en el análisis de datos, visualización y exploración de información.

Capítulo IV

Construcción y evaluación del modelo

Como se mencionó en el Capítulo I, el objetivo de este trabajo es 'Proponer un modelo para la identificación de anomalías en los saldos de las cuentas de depósitos en la información reportada por las instituciones del Sector Financiero Popular y Solidario mediante la aplicación de técnicas de análisis de datos a fin de brindar una herramienta de apoyo al análisis de posibles casos de lavado de activos'. Para cumplir con este objetivo, el cual es a la vez la meta de la tarea de descubrimiento, a continuación, se presenta el desarrollo de la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD) aplicada en este proyecto.

Es importante destacar que la etapa de la metodología KDD denominada 'Comprensión del dominio-negocio' no será abordada en este capítulo, debido a que en el Capítulo I de este trabajo se profundizó en el problema de investigación, sus causas y efectos; así como los objetivos e hipótesis de este trabajo práctico. Además, en el Capítulo II se realizó una revisión inicial de literatura para conocer el estado del arte relacionado al uso de técnicas de análisis de datos para identificar y combatir el lavado de dinero en los sistemas financieros.

Selección de los datos

Identificación de las fuentes de datos

Tomando en consideración los Objetivos del proyecto, para el desarrollo de este trabajo se cuenta con la información de las cuentas de depósito de las entidades del Sector Financiero Popular y Solidario, la cual fue proporcionada por la Superintendencia de Economía Popular y Solidaria, estrictamente para fines analíticos y de investigación académica. Además de la información de depósitos, se cuenta con el catastro histórico de las entidades controladas por la Superintendencia de Economía Popular y Solidaria.

Debido a su carácter de sensibilidad y confidencialidad, esta información se encuentra encriptada, de tal manera que los datos de identificación de las entidades y de los depositantes están codificada.

A continuación, en la Tabla 5 y Tabla 6 se presentan las principales características y condiciones de la información disponible para este trabajo.

Tabla 5

Información de depósitos disponible para la tarea KDD

Concepto	Descripción
Fuente de datos	Sistema de acopio de información, Estructura de Depósitos (D01)
Formato disponible	36 archivos con extensión .zsav (SPSS Statistics comprimido)
Periodo de tiempo	Enero 2018 a diciembre 2020
Número de variables	35
Cantidad de registros	397.174.168
Detalle de datos	Contiene información de los saldos de las cuentas de depósitos reportadas por las entidades a la Superintendencia de Economía Popular y Solidaria, además de la cantidad de depósitos y de retiros, tipo de cuenta de depósito, entre otros.

Tabla 6

Información catastral disponible para la tarea KDD

Concepto	Descripción
Fuente de datos	Sistema de gestión de entidades del sector financiero (GOSF)
Formato disponible	1 archivo con extensión .zsav (SPSS Statistics comprimido)
Periodo de tiempo	Enero 2018 a diciembre 2020
Número de variables	4
Cantidad de registros	36.757
Detalle de datos	Contiene información correspondiente al estado jurídico y segmento de las entidades supervisadas por la Superintendencia de Economía Popular y Solidaria.

Análisis y selección de campos

Debido a la cantidad de información, es necesario que de las bases de datos disponibles se seleccionen únicamente aquellas variables necesarias para aplicar la tarea de descubrimiento. En este sentido, en la Tabla 7 y Tabla 8 se muestra el detalle de las variables existentes en las bases de datos y una descripción conceptual de cada una de ellas.

Tabla 7

Detalle de variables del Sistema de acopio de información - Estructura de Depósitos (D01)

Nombre de la variable	Descripción
NUM_RUC	Es el Número de Registro de Contribuyente de cada entidad del SFPS.
NOM_RAZON_SOCIAL	Corresponde al nombre de cada entidad del SFPS.
FEC_CORTE_DATOS	Fecha contable que corresponde al último día del mes en que la entidad del SFPS reporta la estructura de depósitos (D01).
TIPO_CUENTA	Tipo de obligación que la entidad mantiene con el depositante, p. ej., V = Depósito a la vista.
COD_PROVINCIA	Código de la provincia del Ecuador en dónde se originó la operación de depósito.
PROVINCIA	Nombre de la provincia del Ecuador en dónde se originó la operación de depósito.
COD_CANTON	Código del cantón del Ecuador en dónde se originó la operación de depósito.
CANTON	Nombre del cantón del Ecuador en dónde se originó la operación de depósito.
COD_PARROQUIA	Código de la parroquia del Ecuador en dónde se originó la operación de depósito.
PARROQUIA	Nombre de la parroquia del Ecuador en dónde se originó la operación de depósito.

Nombre de la variable	Descripción
TIPO_OPERACION	Código que determina el estado del depósito, p. ej., VG = Vigente.
BANDA_MADURACION	Corresponde a la cuenta contable que hace referencia al vencimiento del depósito a plazo fijo.
TIPO_IDENTIFICACION	Código del tipo de identificación del depositante, p. ej., C = Cédula.
NUM_IDENTIFICACION	Número de identificación del depositante.
COD_OFICINA	Código de la oficina de la entidad del SFPS en donde se originó la operación de depósito.
NUM_CODIGO_IDENTIFICACION	Código único que otorga cada entidad del SFPS a sus depositantes.
NUM_CUENTA_DOCUMENTO	Número del tipo de cuenta asignado a cada depositante por la entidad del SFPS.
VAL_SALDO	Valor de capital que, a la fecha de corte, registra la operación de depósito.
VAL_INTERES_POR_PAGAR	Valor del interés por pagar que la entidad registra por pagar a la fecha de corte.
FEC_ULTIMA_TRANSACCION	Fecha en la que el depositante efectuó su última transacción.
FEC_EMISION	Fecha en la que se originó la operación de depósito.
FEC_VENCIMIENTO	Fecha en la que finaliza el plazo de la operación de depósito.
NUM_PLAZO	Periodo de tiempo (plazo) para el cuál se estableció la operación de depósito.

Nombre de la variable	Descripción
VAL_SALDO_INICIAL	Es el saldo final del mes inmediato anterior.
VAL_EGRESO_PERIODO	Es el valor total de egresos registrados en la cuenta del depositante.
VAL_INGRESOS_PERIODO	Es el valor total de ingresos registrados en la cuenta del depositante.
NUM_RETIROS_PERIODO	Corresponde al número de transacciones de retiro registrados en la cuenta del depositante.
NUM_DEPOSITOS_PERIODO	Corresponde al número de transacciones de depósito registrados en la cuenta del depositante.
PAIS_NACIMIENTO	Lugar de nacimiento del titular de la cuenta de depósito.
TIPO_DEPOSITO	Código que identifica el tipo de depósito (aplica para depósitos a la vista).
ESTADO_CUENTA	Código que especifica el estado de la cuenta de depósito, p. ej. Abierta.
FEC_CIERRE_CUENTA	Corresponde a la fecha en la cual se cierra la cuenta de depósito.
CAUSAL_CIERRE	Código que especifica el motivo de cierre de la cuenta de depósito.
TARJETA_DEBITO	Determina si el titular de la cuenta de depósito cuenta con una tarjeta de débito.

Tabla 8

Detalle de variables del Sistema de gestión de entidades del sector financiero (GOSF)

Nombre de la variable	Descripción
NUM_RUC	Es el Número de Registro de Contribuyente de cada entidad del SFPS.
FEC_CORTE_DATOS	Fecha correspondiente al último día del mes del catastro del SFPS.
ESTADO_JURIDICO	Es el estado jurídico de la entidad del SFPS en cada fecha de corte.
SEGMENTO	Es el segmento en el cual se encuentra ubicada la entidad del SFPS en cada fecha de corte.

Tomando en consideración el objetivo de la meta KDD y la definición conceptual de las variables disponibles en las bases de datos, del total de variables previamente descritas se seleccionan aquellas que: i) permiten identificar al depositante, ii) permiten evaluar el comportamiento de las cuentas de depósito en términos de movimientos monetarios y cantidad de transacciones; y, iii) actúan como filtros de selección de casos o sirven para resumir la información. En tal sentido, en la Tabla 9 se detallan las variables preliminarmente seleccionadas.

Tabla 9

Variables seleccionadas para el preprocesamiento de la información

Tipo de variable	Fuente	Nombre de la variable	Motivo
Variable de identificación	Base de depósitos	NUM_RUC FEC_CORTE_DATOS NUM_IDENTIFICACION	Estas variables permiten caracterizar e identificar a la operación de depósito en función de la entidad a la que pertenece, el depositante y el periodo de tiempo.
Variable de análisis de comportamiento	Base de depósitos	VAL_SALDO VAL_EGRESO_PERIODO VAL_INGRESOS_PERIODO NUM_RETIROS_PERIODO NUM_DEPOSITOS_PERIODO	Estas variables permiten evaluar el comportamiento de la cuenta de depósitos en términos monetarios y de actividad (depósitos y retiros)
Variable de filtro de casos	Base de depósitos	ESTADO_CUENTA	Esta variable permite identificar si la cuenta de depósito se encuentra vigente o cerrada.
Variable de resumen de casos	Base de depósitos	TIPO_DEPOSITO	Esta variable presenta la información del campo TIPO_CUENTA a un nivel de desagregado.

Tipo de variable	Fuente	Nombre de la variable	Motivo
Variable de identificación y filtro de casos	Base de catastro	NUM_RUC ESTADO_JURIDICO SEGMENTO FECHA_CORTE	Estas variables permiten identificar a la entidad del SFPS, su estado jurídico y segmento a cuál pertenece.

Integración de los datos

En esta actividad se procede a consolidar las bases de datos en un solo conjunto. Para el efecto, se siguen los siguientes pasos:

1. Consolidación en un solo conjunto de datos, de los 36 archivos en formato .szav cuya fuente es el Sistema de acopio de información - Estructura de Depósitos (D01).
2. Incorporación de las variables de la base de datos del Sistema de gestión de entidades del sector financiero (GOSF) al conjunto de datos detallado en el punto anterior.
3. Selección de las variables definidas en la Tabla 9.

Preprocesamiento de los datos

Identificación de casos duplicados

De acuerdo con el Manual Técnico de datos de la estructura de Depósitos - D01 (Superintendencia de Economía Popular y Solidaria, 2022), si existe más de un registro con la misma información en los campos tipo de identificación, identificación, tipo de cuenta, tipo de depósito y número de cuenta – documento, dicho registro es considerado como duplicado.

Tomando en cuenta que en la base de datos se cuenta con información de todas las entidades del SFPS y de varias fechas de corte, además de los campos referidos en el párrafo anterior, se debe incluir a los campos número de RUC y fecha de corte.

Exclusión de registros

Dado que el objetivo de la tarea de descubrimiento es la identificación de anomalías en operaciones de depósito que pertenezcan a entidades con estado jurídico activo y de cuentas de depósito vigentes, en esta actividad se procede a excluir los registros de cuentas cerradas, así como los registros que no pertenecen a entidades con estado jurídico activo.

Ajuste de formato de variables

Algunas variables que son de tipo numérico han sido importadas por la herramienta de análisis en formato de texto, motivo por el que se requiere que dichas variables sean convertidas en tipo numérico, previo a la fase de transformación. Este procedimiento se aplica con las siguientes variables: VAL_SALDO, VAL_EGRESO_PERIODO, VAL_INGRESO_PERIODO, NUM_RETIROS_PERIODO y NUM_DEPOSITOS_PERIODO.

Transformación de los datos

Agregación de casos

Toda cuenta de depósito existente en la base de datos se encuentra vinculada a una persona natural o jurídica, lo cual implica que una persona puede tener más de una cuenta de depósito en una o varias entidades. En este sentido, con el fin de disminuir el número de registros a analizar, se procede a consolidar o resumir la información considerando las variables FEC_CORTE_DATOS y NUM_IDENTIFICACION.

Reducción de dimensiones

Resultado de la aplicación de los pasos anteriores, se cuenta con un conjunto de datos compuesto por 7 variables de las cuales 2 son variables que permiten identificar a la operación de depósito y 5 corresponden a variables que hacen referencia a los movimientos contables en los saldos de las cuentas de depósito.

En este sentido, en esta etapa se procede a realizar una reducción de dimensiones sobre las variables relacionadas con los movimientos contables, mediante el uso del Análisis de Componentes Principales (ACP).

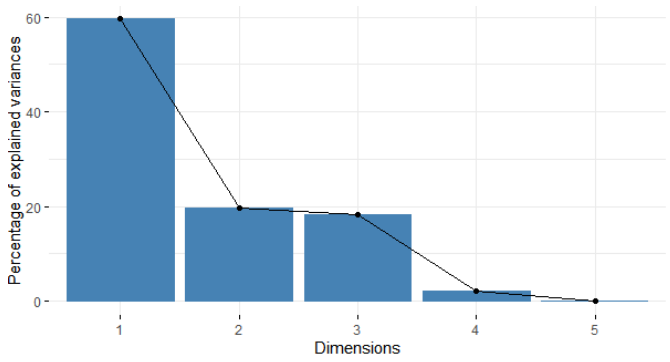
Para llevar a cabo este proceso, se ejecutan las siguientes acciones:

1. Creación de un subconjunto de datos con las variables de tipo numérico.
2. Normalización de las variables y cálculo de las componentes principales.
3. Selección de las componentes principales que en conjunto resuman más del 80% de la varianza del conjunto de datos original.
4. Incorporación de las componentes principales seleccionada en paso anterior al conjunto de datos original.

Para los ensayos realizados en este proyecto, el Análisis de Componentes Principales (ACP) dio como resultado que más del 80% de la varianza de los valores de los elementos del conjunto de datos puede ser explicada a través de tres componentes principales. En la Figura 5 se puede visualizar el porcentaje de la varianza para una entidad seleccionada aleatoriamente.

Figura 5

Porcentaje de varianza explicada por componente principal



Nota. La figura es una captura de pantalla en la que se visualiza que, para el caso de análisis, las componentes principales PC1, PC2 y PC3 explican alrededor del 98% de la varianza. El

código desarrollado en R para la generación del Análisis de Componentes Principales se encuentra en el Apéndice.

Es importante destacar que cada componente principal explica en mayor o menor grado la varianza de las variables del conjunto de datos analizado. Para el caso de ejemplo se puede concluir que la primera componente principal (PC1) está altamente correlacionada con las variables VAL_EGRESO_PERIODO, VAL_INGRESOS_PERIODO y NUM_DEPOSITOS_PERIODO; la segunda componente principal (PC2) está principalmente correlacionada con la variable VAL_SALDO; y, la tercera componente principal (PC3) está altamente correlacionada con la variable NUM_RETIROS_PERIODO. Lo mencionado en este párrafo puede ser observado en la Figura 6.

Figura 6

Matriz de correlación entre variables y componentes principales

	PC1	PC2	PC3
VAL_SALDO	0.2235360	0.85171888	-0.47383160
VAL_EGRESO_PERIODO	0.9872248	-0.04631854	0.04814071
VAL_INGRESOS_PERIODO	0.9879372	-0.03995816	0.04736937
NUM_RETIROS_PERIODO	0.5046255	-0.47845327	-0.70842125
NUM_DEPOSITOS_PERIODO	0.8552755	0.15930872	0.43153607

Nota. La figura es una captura de pantalla en la que se visualiza el valor de correlación entre las componentes principales y las variables de análisis relacionadas a movimientos contables en los saldos de depósito. El código desarrollado en R para la generación de la matriz de correlación se encuentra disponible en el Apéndice.

Selección de registros

Según lo establecido en la Ley Orgánica de Prevención, Detección y Erradicación del Delito de Lavado de Activos y del Financiamiento de Delitos (LOPDEDLAFD) y su reglamento (Asamblea Nacional, 2016), son operaciones o transacciones económicas inusuales,

injustificadas o sospechosas, aquellas que no guarden concordancia con el perfil económico financiero de la persona natural o jurídica, y que no han podido ser sustentadas. Además, establece que los sujetos obligados a reportar, es decir, las instituciones del sistema financiero y de seguros, deben presentar un informe de operaciones y transacciones individuales cuyo valor monetario sea igual o superior a los diez mil dólares de los Estados Unidos de América; así como las diferentes transacciones y operaciones que en conjunto superen o sean igual al valor antes referido, siempre y cuando el beneficiario sea la misma persona y se hayan efectuado en un periodo no mayor a 30 días.

Por lo antes descrito, de la base de depósitos se procede a seleccionar aquellos depositantes que, al menos en una fecha de corte, tengan un saldo de depósito mayor o igual a diez mil dólares o cuyo monto de depósitos o retiros es igual o superior a dicho umbral (diez mil dólares).

Finalmente, debido a que para la identificación de las anomalías en los saldos de depósitos se empleará un análisis de series de tiempo del comportamiento económico de cada depositante, se procede a seleccionar aquellos depositantes que tengan al menos 25 registros, es decir, un historial contable de sus depósitos de al menos 25 meses.

Minería de datos

Selección de la tarea de minería de datos

En términos generales, los procesos de minería contemplan dos tipos de tarea. En primer lugar, tenemos a las tareas de predicción, que pertenecen al grupo de técnicas de aprendizaje supervisado y cuya finalidad es determinar el valor de una variable dependiente, en función a otras variables independientes. Para emplear tareas de medición, es preciso contar con un conjunto de validación o conocer de forma preliminar los valores que puede tomar la variable a predecir.

Por otro lado, están las tareas de descripción que son parte de las técnicas de aprendizaje no supervisado, y se diferencian de las tareas de predicción en que no se conoce de forma preliminar la variable a ser predicha, por lo tanto, lo que se busca es identificar patrones y extraer conocimiento de la base de datos.

Con este antecedente, considerando el objetivo de la meta KDD de este trabajo, se emplearán tareas de minería de datos de tipo descriptiva.

Selección del algoritmo de datos

Tomando en consideración la meta KDD, en este proyecto se emplean dos algoritmos de minería de datos, cuya finalidad es caracterizar a los subconjuntos que conforman la base de datos a analizar y, además, evaluar el comportamiento histórico de los movimientos contables de los depositantes. Para lograr este objetivo, se recurre a las técnicas de análisis de clúster y análisis de anomalías en series temporales.

El análisis de clúster es un conjunto de técnicas que tienen por objeto la clasificación de los individuos y es empleada en escenarios en los cuáles, a priori, se desconoce la pertenencia de los individuos en diferentes grupos. La agrupación de los individuos se basa en la similitud de las propiedades que comparten los elementos agrupados, permitiendo de esta manera describir patrones de similaridad y disimilaridad (Everitt, Landau, Leese, & Stahl, 2011).

Las series temporales son un conjunto de observaciones secuenciales de una variable dispuestas en un orden cronológico, cuyo estudio tiene como finalidad extraer regularidades o patrones relacionados al comportamiento de la variable a través del tiempo. En este sentido, el análisis de las series temporales es un conjunto de técnicas estadísticas que buscan estudiar y modelizar el comportamiento de un fenómeno en el tiempo (Uriel, 1985). De lo anterior se deriva que la detección de anomalías es esencial en el análisis de series de tiempo, ya que a

través de esta se pueden identificar comportamientos anómalos tales como: cambios de tendencia, sucesos inesperados o eventos anormales.

En función a lo descrito en líneas anteriores y tomando en consideración (i) la disponibilidad de información para la ejecución de este trabajo, (ii) los resultados obtenidos en el proceso de revisión inicial de la literatura y (iii) los objetivos del presente trabajo, se empleó el algoritmo k-means para la tarea de clasificación (clustering) y el análisis de los residuales de las series de tiempo para la identificación de anomalías.

K-means fue empleado debido a que el conjunto de datos disponible para ejecutar la tarea de minería, no contaba con registros previamente etiquetados en donde se hayan identificado los datos con aparentes anomalías. Además, k-means es un algoritmo que permite identificar y clasificar observaciones por patrones de comportamiento (conductuales) así como la identificación de valores atípicos (anomalías o desviaciones). En complemento al algoritmo k-means, tomando en consideración que el conjunto de datos disponible cuenta con información de los depositantes en diferentes momentos del tiempo (fechas de corte), se empleó el análisis de las residuales de las series de tiempo, ya que de esta manera se pudo modelar el comportamiento de los depositantes e identificar anomalías a partir de desviaciones en la tendencia del comportamiento de cada sujeto, es decir, la identificación de datos atípicos en la evolución de los saldos de las cuentas de depósito, que es el fin de este trabajo.

A continuación, se presenta una breve explicación de estos algoritmos.

Algoritmo k-means

Es un algoritmo de clasificación de tipo no supervisado que agrupa los elementos de un conjunto de datos en k grupos. La tarea de agrupación se basa en la distancia mínima entre cada elemento y el centroide o clúster del grupo a cuál pertenece (Everitt, Landau, Leese, & Stahl, 2011). Por lo antes descrito, el algoritmo k-means tiene como finalidad identificar la

solución a un problema de optimización, basado en el principio de minimización de la suma de las distancias entre cada elemento y el centroide o clúster de su grupo. Esta tarea, consta de tres pasos:

1. **Inicialización:** consiste en el establecimiento de k centroides, los cuales son seleccionados aleatoriamente del conjunto de elementos a analizar.
2. **Asignación de elementos a los centroides:** en esta instancia cada elemento del conjunto de datos es asignado a su centroide más cercano.
3. **Actualización de centroides:** es un proceso en el cual se actualiza la posición del centroide de cada grupo o clúster a partir de la posición promedio de los elementos que pertenecen al grupo.

Análisis de los residuales de las series de tiempo

La identificación de anomalías es una tarea crítica en muchas disciplinas, en especial, en el análisis de series temporales. Partiendo del concepto de que una anomalía es un suceso inesperado o comportamiento anormal de una variable, la identificación de dichas anomalías se la puede abordar considerando las siguientes tareas: i) Cálculo de residuales en series de tiempo y ii) Detección de anomalías en los residuales de las series de tiempo (Uriel, 1985).

i. Cálculo de residuales en series de tiempo

Los residuales son los elementos remanentes o restantes de una serie de tiempo una vez que se han eliminado los componentes estacionales y de tendencia de la misma serie de tiempo. A este proceso generalmente se lo conoce como descomposición estacional, tiene como finalidad resaltar las anomalías y se lo puede lograr a través de las siguientes técnicas: a) Descomposición estacional de series temporales por Loess[1] y b) Descomposición estacional de series temporales por mediana.

[1] Suavizado de diagramas de dispersión estimado localmente

ii. Detección de residuales en las series de tiempo

Una vez calculados los residuales en la serie tiempo, la siguiente actividad es la identificación de anomalías en dichos residuales, es decir, la detección de elementos que distorsionan la distribución de la serie temporal. Para abordar esta tarea se puede recurrir a dos métodos: a) Rango de cuartil interno y b) Prueba de desviación extrema generalizada.

Empleo del algoritmo de datos

El algoritmo k-means y el análisis de los residuales en las series de tiempo, se aplicó a través del software de procesamiento estadístico R.

R, cuenta con una librería completa para la aplicación de diferentes técnicas de análisis de datos, en tal sentido, en la Figura 7 se presenta el código desarrollado para ejecutar el análisis de clúster a través del algoritmo k-means, mientras que en la Figura 9 se visualiza el código ejecutado para realizar el análisis de los residuales en las series de tiempo.

Figura 7

Captura de pantalla del código en R para el análisis de clúster

```
# ============================================================
# ANÁLISIS DE CLUSTER
# ============================================================
#Cálculo tamaño de muestra
z=1.960; p=0.5; q=1-p; N=nrow(depositos_pca_res); e=0.05 #Parámetros: Nivel de confianza 95% y error máximo 5%
n=round((z^2*N*p*q)/(e^2*(N-1)+z^2*p*q),0) #Tamaño muestra

#Extrae muestra para estimar el número de clústers
depositos_pca_res_muestra <- depositos_pca_res %>%
  sample_n(size = n, replace = FALSE)

#Cálculo número óptimo de clústers
fviz_nbclust(X=depositos_pca_res_muestra[6:8], FUNcluster = kmeans, method = "wss", k.max = 15,
             diss = get_dist(depositos_pca_res_muestra, method = "euclidean"), nstar = 50)

#Generación de cluster: k-means
cluster <- kmeans(depositos_pca_res[6:8],6)
depositos_pca_cluster <- cbind(depositos_pca_res, cluster$cluster)
depositos_pca_cluster_acum <- cbind(depositos_agg[1:2],depositos_pca_cluster)
#ggplot(depositos_pca_cluster,aes(PC1,PC2)) + geom_point(colour=cluster$cluster) + geom_text(aes(label=row.names(depositc

#with(depositos_pca_cluster, plot3d(PC1,PC2,PC3, col = depositos_pca_cluster$cluster, size = 0.25, type = "s"))
plot_ly(x=depositos_pca_cluster_acum$PC1, y=depositos_pca_cluster_acum$PC2, z=depositos_pca_cluster_acum$PC3, size = 0.1

#Almacena resultados
write_xlsx(depositos_pca_cluster_acum, "C://Users//Jonathan//Downloads//cluster_kmeans.xlsx")
```

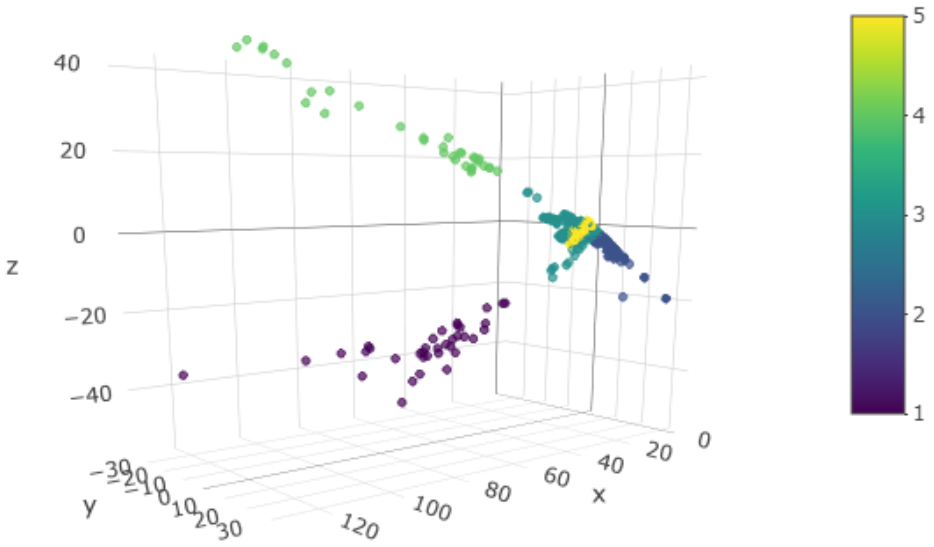
Nota. La figura es una captura de pantalla del código en R desarrollado para aplicar el análisis de clúster a la base de datos objeto de la tarea de minería de datos. El código completo, incluyendo la fase de preprocesamiento, se encuentra disponible en el Apéndice.

Nótese que en la Figura 7, como paso previo al análisis de clústeres, se realiza un muestreo aleatorio simple. Esto se realiza con la finalidad de determinar el número óptimo de clústeres, el cual es un parámetro necesario que debe ser proporcionado por el usuario.

Como resultado de la aplicación del algoritmo k-means, para el caso de análisis, los elementos del conjunto de datos son agrupados en 5 grupos como se puede observar en la Figura 8.

Figura 8

Clústeres generados mediante el algoritmo k-means

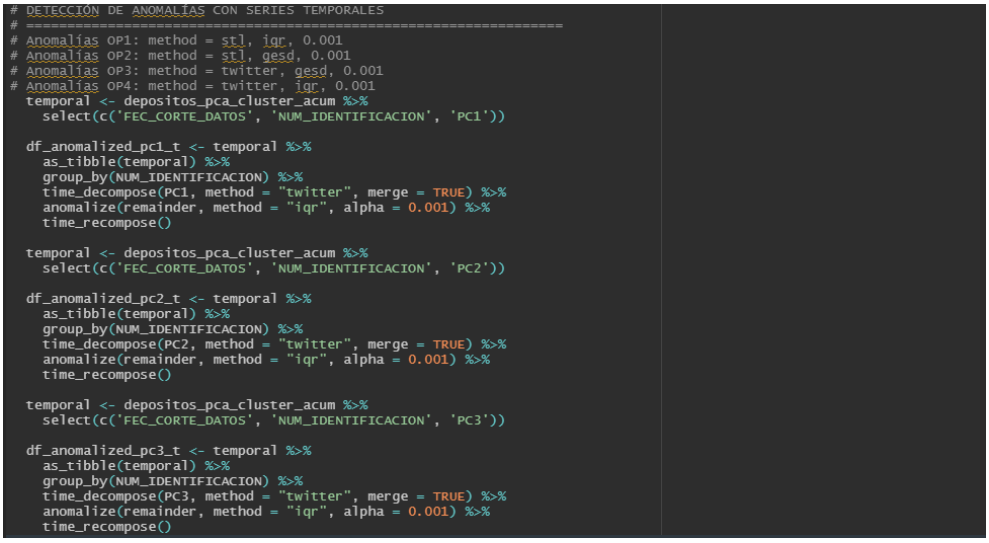


Nota. La figura es una representación gráfica de los clústeres generados a través de la aplicación del algoritmo k-means a las componentes principales, en donde: x = primera componente principal (PC1), y = segunda componente principal (PC2), z = tercera componente principal (PC3). Cada color es un clúster diferente. El código desarrollado en R para la generación de este gráfico se encuentra en el Apéndice.

Con relación al análisis de los residuales de las series de tiempo, en la Figura 9 se puede apreciar el código empleado a través del software R para su ejecución.

Figura 9

Captura de pantalla del código en R para el análisis de residuales en series de tiempo



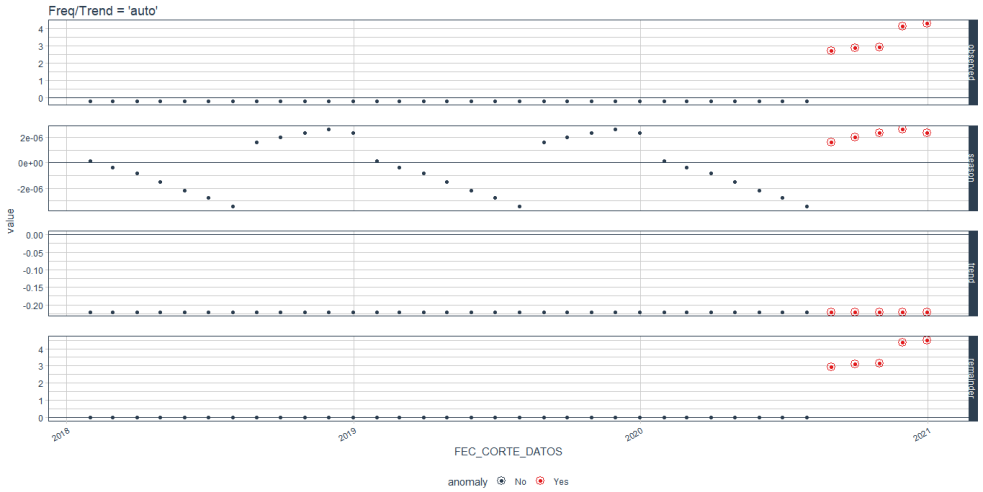
```r
# DETECCIÓN DE ANOMALÍAS CON SERIES TEMPORALES
# ============================================================
# Anomalías OP1: method = stl, iqr, 0.001
# Anomalías OP2: method = stl, gesd, 0.001
# Anomalías OP3: method = twitter, gesd, 0.001
# Anomalías OP4: method = twitter, iqr, 0.001
  temporal <- depositos_pca_cluster_acum %>%
    select(c('FEC_CORTE_DATOS', 'NUM_IDENTIFICACION', 'PC1'))

  df_anomalized_pc1_t <- temporal %>%
    as_tibble(temporal) %>%
    group_by(NUM_IDENTIFICACION) %>%
    time_decompose(PC1, method = "twitter", merge = TRUE) %>%
    anomalize(remainder, method = "iqr", alpha = 0.001) %>%
    time_recompose()

  temporal <- depositos_pca_cluster_acum %>%
    select(c('FEC_CORTE_DATOS', 'NUM_IDENTIFICACION', 'PC2'))

  df_anomalized_pc2_t <- temporal %>%
    as_tibble(temporal) %>%
    group_by(NUM_IDENTIFICACION) %>%
    time_decompose(PC2, method = "twitter", merge = TRUE) %>%
    anomalize(remainder, method = "iqr", alpha = 0.001) %>%
    time_recompose()

  temporal <- depositos_pca_cluster_acum %>%
    select(c('FEC_CORTE_DATOS', 'NUM_IDENTIFICACION', 'PC3'))

  df_anomalized_pc3_t <- temporal %>%
    as_tibble(temporal) %>%
    group_by(NUM_IDENTIFICACION) %>%
    time_decompose(PC3, method = "twitter", merge = TRUE) %>%
    anomalize(remainder, method = "iqr", alpha = 0.001) %>%
    time_recompose()
```

Nota. La figura es una captura de pantalla del código en R desarrollado para ejecutar el análisis de los residuales en las series de tiempo de los saldos y movimientos contables de cada depositante, a través de sus componentes principales. El código completo, incluyendo la fase de preprocesamiento, se encuentra disponible en el Apéndice.

Un aspecto que destacar en el código visualizado en la Figura 9, es que el análisis de residuales en las series de tiempo se aplica a cada uno de los depositantes, es decir, se evalúa el comportamiento en el tiempo de cada depositante y a partir de dicho análisis se establecen las anomalías, como se puede apreciar en la Figura 10.

Figura 10

Identificación de anomalías por descomposición estacional de series de tiempo mediante Loess



Nota. La figura es una representación gráfica de la descomposición estacional de series de tiempo por Loess. El código desarrollado en R para la generación de este gráfico se encuentra en el Apéndice.

Otro aspecto para destacar es que, tanto en el análisis de clúster como en el análisis de las residuales en series de tiempo, los algoritmos son aplicados a las componentes principales, mismas que fueron determinadas como parte del proceso de reducción de dimensionalidad del conjunto de datos realizada en la fase de Transformación de los datos. Se destaca este particular debido a que el análisis de componentes principales (ACP) no solo permitió disminuir la dimensionalidad de la base de datos objeto de la meta KDD, sino que también implicó la normalización de los valores de los elementos de dicho conjunto de datos, con lo cual se optimiza el proceso de identificación de anomalías y se disminuye la posibilidad de identificar erróneamente casos que no son anomalías.

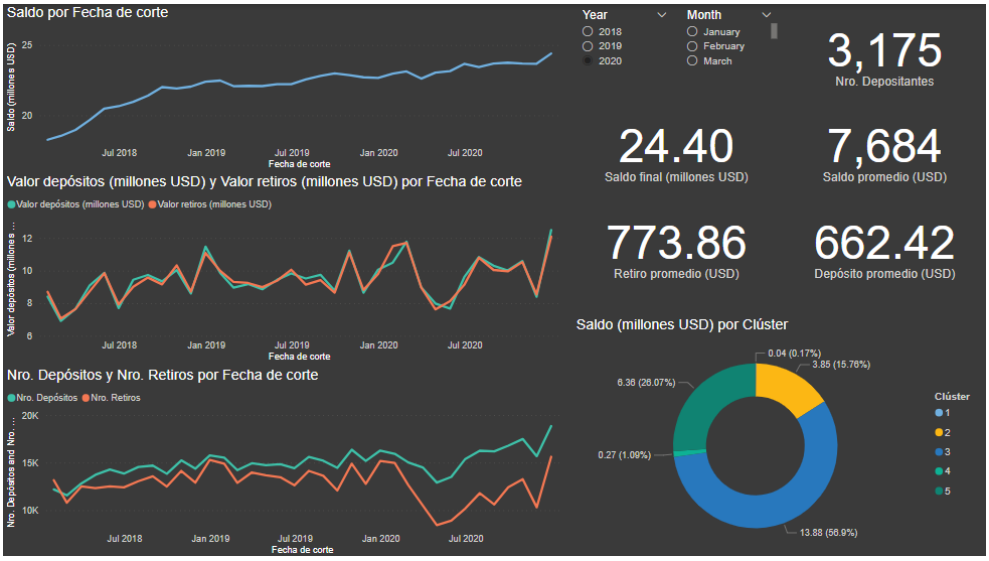
Presentación, interpretación y evaluación de resultados

Los resultados obtenidos del análisis de clúster y análisis de residuales en las series de tiempo son presentados a través de un informe construido con la herramienta de inteligencia de negocios denominada Power BI. Este informe contiene las siguientes pestañas:

1. **Perfil general:** en esta pestaña se presentan indicadores clave tales como la evolución del saldo de los depósitos, la fluctuación en el tiempo de los depósitos y retiros expresados en monto y cantidad, así como también un perfil general de los grupos de depositantes por clúster. Este reporte puede visualizarse en la Figura 11.

Figura 11

Captura de pantalla de la pestaña 'Perfil General' del reporte de identificación de anomalías



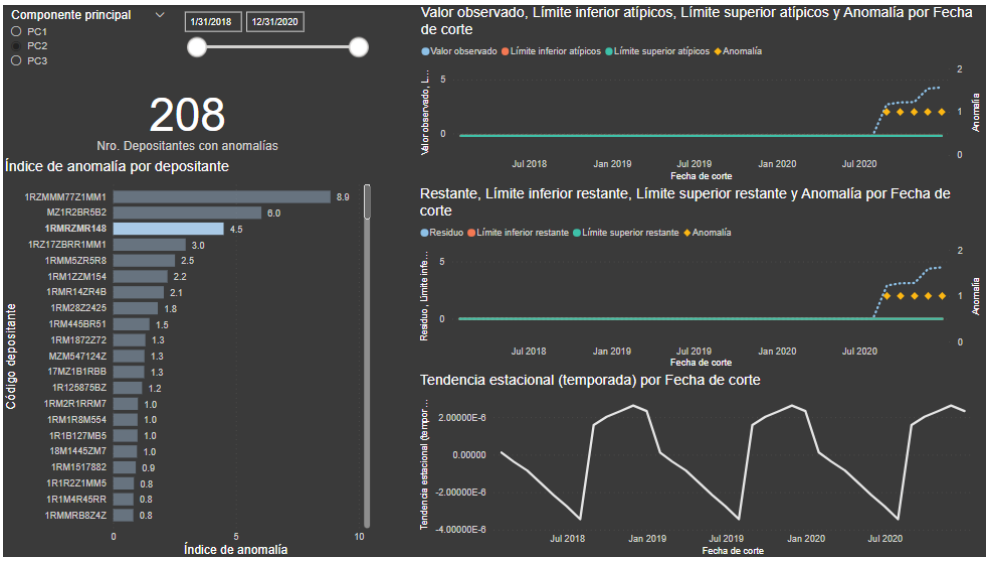
Nota. La figura es una captura de pantalla de la pestaña Perfil General del reporte de identificación de anomalías en los depósitos, el cual ha sido desarrollado en Power BI.

2. **Depositantes con operaciones anómalas:** en esta sección se visualizan todos aquellos depositantes que, de acuerdo con el análisis de residuales en la serie de tiempo, tienen al menos una instancia atípica en el periodo de análisis en una o más

de las componentes principales analizadas. Por defecto, la pestaña presenta en orden descendente el código del depositante en función de las anomalías en su saldo de depósitos o movimientos (depósitos y retiros). Además, se visualizan los componentes estacionales y componentes de tendencia a partir de los cuáles se determina la existencia de una anomalía. En la Figura 12 se muestra una captura de pantalla de este reporte.

Figura 12

Captura de pantalla de la pestaña 'Depositantes con operaciones anómalas' del reporte de identificación de anomalías

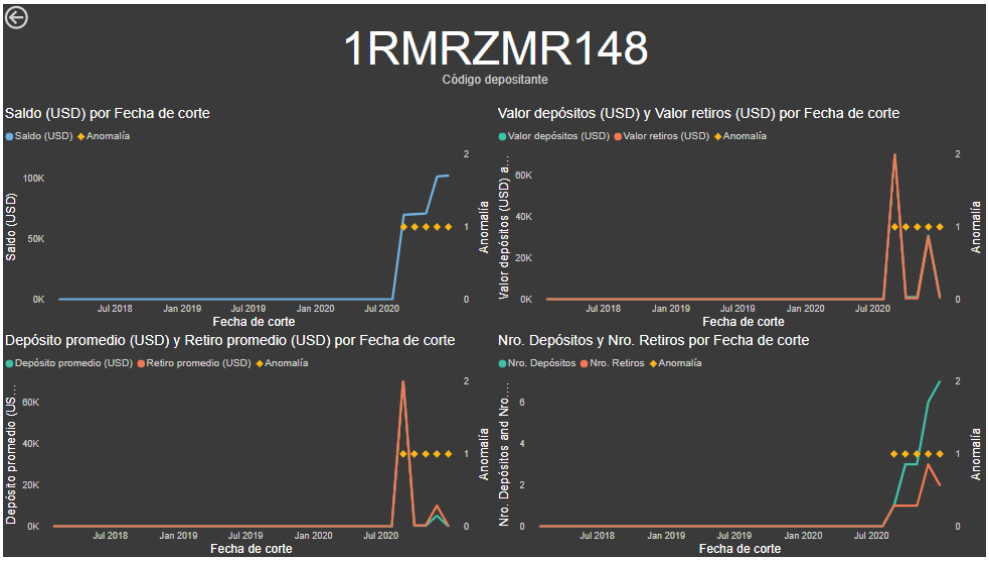


Nota. La figura es una captura de pantalla de la pestaña Depositantes con operaciones anómalas del reporte de identificación de anomalías en los depósitos, el cual ha sido desarrollado en Power BI. El índice de anomalía se calcula a partir de la diferencia entre el valor observado y los límites superior e inferior del valor recompuesto, a mayor distancia del valor observado por encima del límite, mayor es el valor del índice de anomalía.

3. **Detalle depositante:** en esta pestaña se puede revisar los motivos por los cuáles se ha identificado que el depositante tiene anomalías en su saldo de depósitos o movimientos (depósitos y retiros en monto y cantidad).

Figura 13

Captura de pantalla de la pestaña Detalle depositante



Nota. La figura es una captura de pantalla de la pestaña Detalle depositante del reporte de identificación de anomalías en los depósitos, el cual ha sido desarrollado en Power BI.

A partir del ejemplo ilustrado en las imágenes tomadas del Reporte de identificación de anomalías, se puede llegar a las conclusiones:

- La entidad objeto de análisis tiene más de tres mil depositantes que en el periodo de tiempo analizado (enero 2018 a diciembre 2020), presentaron en al menos un mes, un saldo de depósitos o movimientos (depósitos o retiros) por un monto mayor o igual a diez mil dólares de Estados Unidos de América.
- De acuerdo con el saldo de depósitos o movimientos (depósitos o retiros), la entidad tiene cinco diferentes grupos de depositantes.

- 208 depositantes, es decir, el 7% del total de los depositantes de la entidad, presentan anomalías en su saldo de depósitos o movimientos (depósitos o retiros).
- Una de las anomalías más significativas según la segunda componente principal (PC2), le corresponde al código de depositante 1RMRZMR148, mismo que presentó una variación significativa en su saldo de depósitos, monto y cantidad de retiros y depósitos, entre los meses de agosto y diciembre del año 2020.

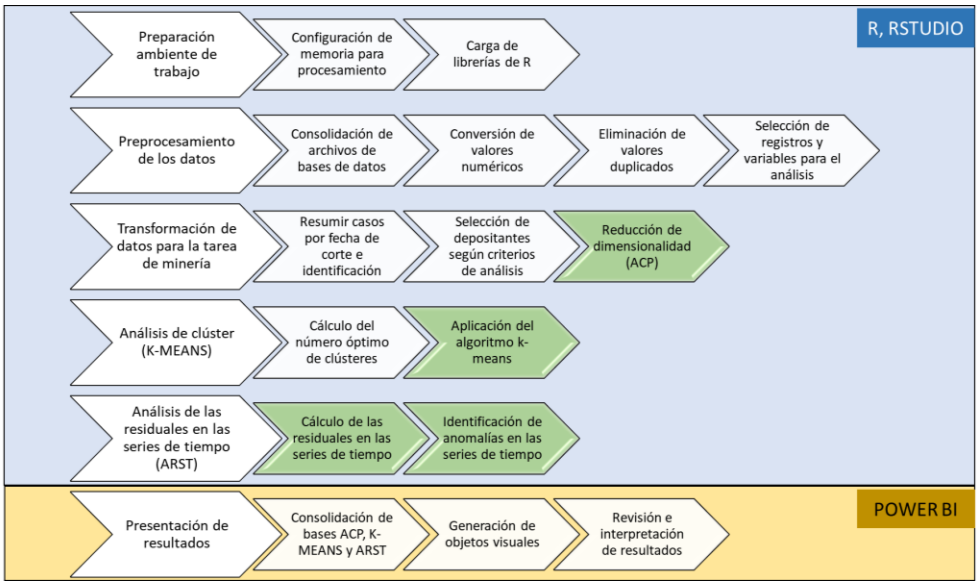
El análisis y las conclusiones determinadas a partir del ejemplo ilustrado en la Figura 11, Figura 12 y Figura 13, se puede realizar para el resto de los depositantes en los cuáles se ha identificado anomalías. Es así como, de esta manera, el modelo presenta el listado de depositantes con anomalías en sus saldos o movimientos de depósitos, y a través del índice de anomalía disponible en el Reporte de identificación de anomalías, el área de supervisión y monitoreo puede concentrar sus esfuerzos en analizar con un mayor nivel de profundidad los casos de anomalías más significativas, contando de esta forma con una herramienta adicional para la prevención de lavado de dinero.

Esquema gráfico del modelo desarrollado

Todas las actividades desarrolladas en el presente trabajo que tuvo como finalidad la generación de una propuesta de modelo de identificación de anomalías en los saldos de las cuentas de depósitos de las entidades del Sector Financiero Popular y Solidario se visualiza gráficamente a manera de esquema conceptual en la Figura 14.

Figura 14

Esquema conceptual del modelo de identificación de anomalías



Nota. La figura es representación gráfica del esquema conceptual de la propuesta de modelo de identificación de anomalías en los saldos de las cuentas de depósito de las entidades del Sector Financiero Popular y Solidario.

En este punto, es preciso mencionar que la propuesta de modelo presentada en este trabajo puede aplicarse a una entidad (como el caso que se ilustró previamente), a un grupo de entidades o al sector financiero popular y solidario en su conjunto, según los requerimientos de análisis y las capacidades de procesamiento de información disponibles.

Capítulo V

Conclusiones, recomendaciones y trabajos futuros

Una vez concluido el desarrollo de la ‘Propuesta de modelo para la identificación de anomalías en los saldos de las cuentas de depósitos del Sector Financiero Popular y Solidario del Ecuador’, a continuación, se procede a detallar las conclusiones, recomendaciones y trabajos futuros de este proyecto.

Conclusiones

- El lavado de dinero es un proceso a través del cual se busca dar apariencia lícita a los recursos financieros provenientes de actividades ilícitas, siendo el sistema financiero uno de los medios comúnmente utilizados para este fin; de ahí radica la importancia de que el Organismo de Control del Sector Financiero Popular y Solidario de Ecuador disponga de un modelo de identificación de anomalías en los saldos de las cuentas de depósito que le reportan sus entidades supervisadas, de tal manera que cuente con una herramienta adicional de apoyo para el análisis de potenciales casos de lavado de activos.
- La identificación del uso del Sector Financiero Popular y Solidario para el cometimiento del proceso de lavado de dinero nace en primera instancia de la identificación de operaciones anómalas, entre ellas, la detección de comportamientos inusuales en los saldos de las cuentas de depósito de los depositantes. De esto se deriva la importancia de contar con herramientas y modelos que permitan identificar desviaciones en el perfil y comportamiento económico de los depositantes.
- El uso de técnicas de análisis de datos soportado en las tecnologías de información y comunicaciones (TICs), facilita la tarea de identificación de anomalías en bases de datos con una gran cantidad de registros, disminuyendo el tiempo requerido para el

procesamiento de la información, beneficiando a los procesos analíticos de monitoreo y supervisión.

- Existen múltiples técnicas y algoritmos que permiten el procesamiento de la información y a partir de ello identificar anomalías en los registros de las bases de datos, sin embargo, tomando en cuenta la información disponible para la ejecución de este trabajo, se consideró como idóneo el uso de técnicas con un enfoque descriptivo, tales como: el análisis de componentes principales para la reducción de la dimensionalidad, la clasificación de los elementos a través del algoritmo k-means y la identificación de anomalías a través de un análisis de los residuales en las series de tiempo.
- En la actualidad, existen múltiples metodologías que permiten implementar proyectos de minería de datos, entre las que destaca la metodología de 'Descubrimiento de Conocimiento en Bases de Datos' comúnmente conocida como KDD. KDD fue el marco referencial empleado en la ejecución del presente trabajo debido a la facilidad de iterar entre las diferentes fases de la metodología, contemplando desde la comprensión de las necesidades del negocio hasta la interpretación y análisis de los resultados obtenidos.
- Es importante que los resultados de los modelos analíticos desarrollados sean de fácil comprensión para el usuario final, de ahí se origina la necesidad del empleo de herramientas de inteligencia de negocios para que a través elementos visuales permitan comprender el funcionamiento de los modelos analíticos y entender de forma sencilla los resultados obtenidos.

Recomendaciones

- Se recomienda al Organismo de Control del Sector Financiero Popular y Solidario tomar en consideración la propuesta de modelo presentada en este trabajo para la identificación de anomalías en los saldos de las cuentas de depósito de las entidades bajo su supervisión, a fin de fortalecer y apoyar el proceso analítico de detección de anomalías ante posibles casos de lavado de dinero.
- Así también, se recomienda a las entidades que forman parte del Sector Financiero Popular y Solidario y que aún no disponen de un modelo propio de identificación de anomalías en sus captaciones (depósitos), tomen como marco de referencia inicial a la propuesta de modelo presentada en este trabajo, el cual les permitirá no solo identificar anomalías sino también conocer más a detalle el comportamiento y perfil económico de sus depositantes a través del tiempo.

Trabajos futuros

Dado que el presente trabajo se centra en la presentación de una propuesta para la identificación de anomalías en las cuentas de depósito de las entidades del Sector Financiero Popular y Solidario, y que por tanto no es concluyente en la determinación de si las anomalías identificadas tienen relación con casos de lavado de dinero, a futuro podría resultar interesante:

- Incorporar al modelo propuesto otras fuentes de información tales como la información de cartera de crédito de las entidades del Sector Financiero Popular y Solidario y el perfil sociodemográfico de sus depositantes y sujetos de crédito. Esto con la finalidad de evaluar si con dicha información adicional, se identifican nuevos casos de depositantes con anomalías en sus cuentas.
- Además, sería interesante determinar cuántos de los casos identificados como depositantes con anomalías en sus cuentas a través del modelo propuesto en este trabajo, también han sido identificados como depositantes inmersos en actividades de lavado de dinero por parte del Organismo de Control del Sector Financiero Popular y Solidario y otras Carteras de Estado habilitadas para investigar este tipo de ilícitos.

Bibliografía

Asamblea Nacional. (21 de Julio de 2016). Ley Orgánica de Prevención, Detección y Erradicación del Delito de Lavado de Activos y del Financiamiento de Delitos. Quito.

Asamblea Nacional Constituyente. (2008). *Constitución de la República del Ecuador.* Recuperado el 12 de Febrero de 2019, de https://www.asambleanacional.gob.ec/sites/default/files/documents/old/constitucion_de_bolsillo.pdf

Asamblea Nacional Constituyente. (14 de April de 2011). *Asamblea Nacional Constituyente.* Obtenido de https://www.seps.gob.ec/documents/20181/25522/LEY%20ORGANICA%20DE%20ECONOMIA%20POPULAR%20Y%20SOLIDARIA%20actualizada%20noviembre%202018.pdf/66b23eef-8b87-4e3a-b0ba-194c2017e69a

Asamblea Nacional Constituyente. (2011). *Ley Orgánica de Economía Popular y Solidaria.* Quito.

Azevedo, A. (2008). KDD, semma and CRISP-DM: A parallel overview. *International Association for Development of the Information Society*, 182-185.

Basel Institute on Governance. (2019). *International Centre for Asset Recovery*. Recuperado el Febrero de 2019, de https://index.baselgovernance.org/ranking

Bernal, X., & Sares, J. (2019). *Universidad de Cuenca.* Obtenido de http://dspace.ucuenca.edu.ec/bitstream/123456789/31743/1/Trabajo%20de%20Titulaci%C3%B3n.pdf

Bouazza, I., Ameur, E. B., & Ameur, F. (2019). Datamining for Fraud Detecting, State of the Art. *Springer*, 213-214.

Camana, R. (July de 2016). *Revista Tecnológica ESPOL - RTE.* Obtenido de http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/464/338

Cao, Longbing. (2018). Data Science Applications. *Data Science Thinking*, 263-292.

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

Echegoyen, G. (2003). *Registros administrativos, calidad de los datos y credibilidad pública.*

Everitt, B., Landau, S., Leese, M., & Stahl, D. (7 de January de 2011). Cluster Analysis.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *AAAI*, 82-88.

FLACSO. (2015). Lavado de activos. *Perfil criminológico*(15), 15.

Guevara, J., García, O., & Granados, O. (2020). Machine Learning Methodologies Against Money Laundering in Non-Banking Correspondents. *Springer*, 85-86.

Haugeland, J. (1988). *La Inteligencia Artificial.*

Hernández Martínez, F. (2019). *Universidad EIA.* Obtenido de https://core.ac.uk/download/pdf/232126815.pdf

Huo, L., Wang, T., & Liu, Y. (2013). Data Mining Techniques Applied in the Financial Industry. *The 19th International Conference on Industrial Engineering and Engineering Management*, 1237-1243.

Jagadish, H., Gehrke, J., Labrindis, A., Papakonstantinou, Y., Patel, J., Ramarcrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *ACM Digital Library*, 86-94.

Junta de Política y Regulación Monetaria y Financiera. (2014). Norma para la prevención de lavado de activos y financiamiento de delitos incluido el terrorismo en las entidades financieras de la economía popular y solidaria. Quito.

Kashyap, P. (2018). Technology Stack for Machine Learning and Associated Technologies. *Machine Learning for Decision Makers*(858), 137-187.

Keshav Palshikar, G. (2014). Detecting Frauds and Money Laundering: A Tutorial. *International Conference on Big Data Analytics*, 145-160.

Keshav, G. (2014). Detecting Frauds and Money Laundering: A Tutorial. *International Conference on Big Data Analytics*, 145-160.

Kotu, V., & Deshpande, B. (2019). *Data Science: Concepts and Practice.* Chennai: Morgan Kaufmann.

Kurgan, L., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. *The Knowledge Engineering Review*, 1-24.

Mejía Vanegas, H. R. (June de 2018). *Pontificia Universidad Católica del Ecuador.* Obtenido de https://repositorio.pucesa.edu.ec/bitstream/123456789/2435/1/76712.pdf

Montes, A. (2014). *El sector financiero y el lavado de dinero.* Quipukamayoc.

Namakforoosh, M. (2000). *Metodología de la investigación.* México: Limusa.

Organización de Estados Americanos. (2018). *OEA.* Obtenido de https://www.oas.org/es/ssm/ddot/publicaciones/LIBRO%20OEA%20LAVADO%20ACTIVOS%202018_4%20DIGITAL.pdf

Pérez, C. (2007). *Minería de datos: Técnicas y herramientas.* Madrid: 2008.

Pesántez Coyago, M. E. (December de 2020). *Universidad del Azuay.* Obtenido de http://dspace.uazuay.edu.ec/bitstream/datos/10496/1/16085.pdf

Pesántez Coyago, M. E. (December de 2020). *Universidad del Azuay.* Obtenido de http://dspace.uazuay.edu.ec/bitstream/datos/10496/1/16085.pdf

Rojas Sánchez, R. (Septiembre de 2018). *Universidad Católica de Santiago de Guayaquil.* Obtenido de http://201.159.223.180/bitstream/3317/11566/1/T-UCSG-PRE-ECO-CICA-368.pdf

Rojas, H. (2016). *Métodos y enfoques en la investigación cualitativa.* Universidad Industrial de Santander.

Superintendencia de Economía Popular y Solidaria. (2018). *Informe de rendición de cuentas (Preliminar).* Quito.

Superintendencia de Economía Popular y Solidaria. (2019). Estatuto Orgánico de Gestión Organizacional por Procesos. Quito.

Superintendencia de Economía Popular y Solidaria. (2020). *Superintendencia de Economía Popular y Solidaria.* Obtenido de https://www.seps.gob.ec/documents/20181/1026889/Rendici%C3%B3n+Cuentas+2020.pdf/4cc3c161-e0eb-482d-bd84-a8e6edd5bdda?version=1.0

Superintendencia de Economía Popular y Solidaria. (2020). *Superintendencia de Economía Popular y Solidaria.* Obtenido de https://www.seps.gob.ec/documents/20181/1026889/Rendici%C3%B3n+Cuentas+2020.pdf/4cc3c161-e0eb-482d-bd84-a8e6edd5bdda?version=1.0

Superintendencia de Economía Popular y Solidaria. (9 de January de 2022). *Manual Técnico de Datos - Estructura de depósitos.* Obtenido de SEPS:

https://www.seps.gob.ec/documents/20181/374891/Manual+de+Dep%C3%B3sitos_V5.0_02abr2020_c.pdf/5fe8504b-248d-43ce-b30f-48804997f7bf

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., & Hidalgo-Troya, A. y. (2016). *El proceso de descubrimiento de conocimiento en bases de datos.* Bogotá: Ediciones Universidad Cooperativa de Colombia.

Unidad de Análisis Financiero. (21 de May de 2021). *Unidad de Análisis Financiero.* Obtenido de https://www.uafe.gob.ec/rendicion-de-cuentas-2020/

Unidad de Información y Análisis Financiero. (2014). *La dimensión económica del lavado de activos.* Bogotá.

United Nations Office on Drugs and Crime. (2019). *Money-Laundering and Globalization.*

Uriel, E. (1985). *Análisis de series temporales.* Paraninfo.

Vaishnavi, V. (2007). *Design Science Research Methods and Patterns: Innovating Information.* New York: Auerbach Publications.

Vieria, L. (2009). *Introducción a la minería de datos.* Río de Janerio: E-papers.

Wei, W., Jin, j. L., Longbing, C. Y., & Chen, J. (2012). Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 449–475.

Xuan, L., Pengzhu, Z., & Zeng, D. (2008). Sequence Matching for Suspicious Activity Detection in Anti-Money Laundering. *Lecture Notes in Computer Science*, 50-61.

Apéndices

Apéndice A. Código de R desarrollado para la identificación de anomalías en los saldos de las cuentas de depósito