

**SEGMENTACIÓN DE CLIENTES Y DETECCIÓN DE OPERACIONES
INUSUALES CLASIFICADOS EN GRADOS DE RIESGO PARA LA
PREVENCIÓN DE LAVADO DE DINERO CON DATOS DE UNA INSTITUCIÓN
FINANCIERA EN MÉXICO A 2023**
**CLIENT SEGMENTATION AND DETECTION OF UNUSUAL OPERATIONS CLASSIFIED
INTO RISK DEGREES FOR THE PREVENTION OF MONEY LAUNDERING WITH DATA
FROM A FINANCIAL INSTITUTION IN MEXICO BY 2023**

José Antonio Ortiz Richards¹

Banco Nacional de Comercio Exterior, México

(jortizrichards@gmail.com) (<https://orcid.org/0000-0001-9795-8001>)

Información del manuscrito:

Recibido/Received: 19/11/2024

Revisado/Reviewed: 03/12/2024

Aceptado/Accepted: 10/05/2025

RESUMEN

Palabras clave:

minería de datos, prevención de lavado de dinero, operaciones inusuales, métodos de segmentación.

La investigación desarrolla y valida un método cuantitativo utilizando datos de una institución financiera en México, con el objetivo de obtener un mejor conocimiento de los clientes, detectar el posible uso indebido de la institución en delitos relacionados con la integración y dispersión de recursos financieros ilícitos atendiendo las recomendaciones internacionales establecidas por el Grupo de Acción Financiera Internacional (GAFI) y con las disposiciones en México. Se emplean técnicas de minería de datos, así como instrumentos para recolectar, analizar y utilizar la información sobre las características inherentes y transaccionales de los clientes. Se presenta un análisis estadístico descriptivo y, para lograr una segmentación adecuada, se combinan métodos de clasificación basados en centros móviles y la clasificación jerárquica de Ward, junto con métodos factoriales, lo que permite identificar cambios en patrones de comportamiento de las variables y analizar posibles operaciones inusuales, explicando el grado de riesgo asociado a cada segmento. Los resultados ofrecen una clasificación de riesgo medio y alto, contrastando con el modelo de puntaje actualmente utilizado, que clasifica a los clientes como de riesgo bajo. Además, este enfoque facilita la sospecha de operaciones inusuales reduciendo el número de alertas falsas. Uno de los aportes que ofrece esta investigación es incorporar variables cualitativas para la segmentación adaptada al contexto de México considerando mejores prácticas en Colombia y de GAFI.

ABSTRACT

Keywords:

The research develops and validates a quantitative method using data from a financial institution in Mexico, with the objective of obtaining a better understanding of the clients, detecting the

¹ Autor de correspondencia.

data mining, money laundering prevention, unusual operations, cluster formation methods.

possible misuse of the institution in crimes related to the integration and dispersion of illicit financial resources, taking the international recommendations established by the Financial Action Task Force (FATF) and with the provisions in Mexico. Data mining techniques are used, as well as instruments to collect, analyze and use information on the inherent and transactional characteristics of customers. A descriptive statistical analysis is presented and, to achieve adequate segmentation, classification methods based on mobile centers and Ward's hierarchical classification are combined, along with factorial methods, which allows for the identification of changes in behavioral patterns of the variables and the analysis of possible unusual operations, explaining the degree of risk associated with each segment. The results offer a classification of medium and high risk, contrasting with the scoring model currently used, which classifies customers as low risk. In addition, this approach facilitates the suspicion of unusual operations by reducing the number of false alerts. One of the contributions this research offers is the incorporation of qualitative variables for segmentation adapted to the context of Mexico, considering best practices in Colombia and the FATF.

Introducción

Una de las prioridades internacionales es la lucha para la prevención del lavado de dinero y el financiamiento al terrorismo debido a que representan un alto riesgo para la paz, la seguridad, la estabilidad y el desarrollo económico de los países. El efecto de no gestionarse adecuadamente, por una parte, supone un incremento en el número de víctimas a causa de los delitos cometidos por personas u organizaciones y, por otra parte, de acuerdo con BASEL (2022) los inversionistas y las instituciones financieras pueden decidir retirarse o no iniciar un negocio en un país que es evaluado con alto riesgo en lavado de dinero para evitar la exposición a los riesgos de incumplimiento, operacionales y de reputación.

Existen diversos organismos internacionales para combatirlo, un organismo es el GAFI quien periódicamente revisa las técnicas de lavado de dinero, fortalece sus estándares y monitorea a los países para garantizar que implementen de manera completa y efectiva sus 40 recomendaciones. Los datos del GAFI son una fuente primordial para conocer el compromiso del país. México cumple parcialmente algunas recomendaciones, conforme con el proceso de evaluación llevado a cabo entre 2017 y 2018 (GAFILAT,2023).

Otro Organismo, es BASEL que, entre otras acciones, mide anualmente “el riesgo de blanqueo de capitales y financiación del terrorismo (LA/FT) en jurisdicciones de todo el mundo”(BASEL, 2023, p. 5). El riesgo lo define como la vulnerabilidad al LA/FT y sus capacidades para contrarrestarlo. El índice considera 18 indicadores, los cuales difieren en enfoque y alcance y se encuentran clasificados en cinco dominios. La puntuación se basa en datos de fuentes disponibles públicamente, como el GAFI, Banco Mundial, Foro Económico Mundial y Transparencia Internacional. En 2022 y 2023, México esta evaluado en riesgo medio.

El Gobierno Mexicano derivado de los compromisos internacionales adoptados como integrante del GAFI emite para las instituciones financieras, entre diversas leyes y normas, las disposiciones de carácter general a que se refiere el artículo 115 de la Ley de Instituciones de Crédito para prevenir y detectar estos actos, la cual establece que las “Entidades con base en la medición de los Riesgos que lleven a cabo, deberán clasificar a sus Clientes en diversos Grados de Riesgo que permitan diferenciarlos de forma significativa” (SHCP,2022, p. 67).

Por otra parte, la Comisión Nacional Bancaria y de Valores (CNBV) en su labor de supervisión identifica varios hallazgos recurrentes, entre los que destacan: “El modelo de clasificación de clientes por grado de riesgo no consideran todos los criterios, factores y características inherentes o transaccionales de los clientes” (2020, p. 9).

La institución financiera para dar cumplimiento desarrolla y utiliza un modelo de clasificación por puntos. Desde su aplicación presenta resultados con una concentración de clientes en grado de riesgo bajo, los cuales cuentan con características heterogéneas. Se considera que las limitantes de este modelo son:

- Es estático, por lo que no se ajusta a cambios en el comportamiento del cliente.
- Genera un número elevado de falsos positivos.

Se destaca que el LA/FT se ve beneficiado por el uso de la inteligencia artificial ya que mejora de forma significativa y en tiempo real el procesamiento de grandes volúmenes de datos facilitando la precisión de los resultados para identificar operaciones atípicas (GAFI, 2021, p. 35; Martínez et. al., 2022, p. 78).

GAFI (2021, p. 38) menciona que la prevención del LA/FT se ve favorecida por la segmentación, ya que al formar grupos homogéneos de individuos es posible encontrar asociación entre ellos conociendo su comportamiento habitual y a partir de esto identificar operaciones inusuales, en comparación con su segmento que pueden estar vinculadas a comportamientos sospechosos de delitos asociados con el LA/FT.

Por lo tanto, el objetivo general del presente artículo es utilizar un método de segmentación que permita clasificar y detectar la inusualidad de las operaciones de los clientes considerando datos de 2022 y 2023 para contar con mayores elementos para conocerlo, clasificarlo en un grado de riesgo y en su caso sospechar de LD.

Las preguntas de investigación que se plantean son: ¿Cómo podría la institución financiera en México segmentar a los clientes para lograr conformar grupos homogéneos al interior y heterogéneos entre ellos?; ¿La segmentación de los clientes permite explicar el grado de riesgo que sus características inherentes y transaccionales exponen a la institución financiera en México para ser utilizada en el lavado de dinero?; y ¿La institución financiera en México podría detectar y sospechar de operaciones inusuales de sus clientes que representen riesgo de lavado de dinero con el apoyo la segmentación?.

Para ello, inicialmente se recolecta y analizan diversos artículos, como el publicado por (Jovel, 2020) quien segmenta a los asociados de una cooperativa de ahorros y créditos en Colombia utilizando variables cuantitativas. Se apoya del software R, de la minería de datos Cross Industry Standard Process for Data Mining (CRISP DM) y de tres técnicas de agrupamiento basado en particiones: K-means- PAM, Fuzzy C-Means, y una técnica basada en modelos de densidad: Dbscan. Concluye que se cumple con los objetivos planeados al identificar los mismos grupos por los diferentes métodos aplicados, sin embargo, los asociados del “grupo obtenido por PAM, difieren de los obtenidos por K-means, ya que el primero se basa en el medoide, mientras que el segundo en el centroide” (p. 84).

Castro y Castro (2020) desarrollan una propuesta con el objetivo de “identificar grupos homogéneos y heterogéneos entre segmentos a partir de variables cuantitativas de los asociados en un fondo de empleados en Colombia” (p. 10). Se apoya de la herramienta de Rstudio. Realizan un análisis de componentes principales y concluyen que al aplicar el método de segmentación con el algoritmo k medias en línea con el objeto de estudio el resultado de esta agrupación determinó tres grupos e identifica señales de alerta.

Por otra parte, Perez (2020) presenta una metodología de segmentación para dos entidades financieras colombianas utilizando el algoritmo de K-Means y comparando los resultados en diferentes índices que permitan validar la óptima selección del número de clústeres y la pertenencia a estos. Utiliza las siguientes variables: Frecuencia de transacciones, ciudad de origen, sector o persona natural, años de antigüedad y transacciones, con el apoyo del Software R y el paquete NbClustm. Concluye que se obtiene una segmentación adecuada.

En cuanto antecedentes en México, se encuentra el artículo por parte de Camacho et al., (2021) en la que tratan la detección de LA/FT mediante redes neuronales y un indicador de anomalías. La metodología propuesta involucra el uso de lógica difusa para obtener métricas de riesgo, utiliza cuatro algoritmos no supervisados (Strict Competitive Learning, Self-Organizing-Map, C-Means y Neural Gas) para formar grupos e identificar el de mayor riesgo, y por último aplica un indicador de anomalía para detectar cualquier comportamiento inusual. El modelo logra reducir la tasa de falsos positivos y disminuir los costos de la empresa.

Se puede observar que los artículos se desarrollaron principalmente en Colombia con variables cuantitativas diferentes a la del presente artículo, esto debido a los

requerimientos de las entidades regulatorias; la mayoría de las investigaciones precedentes utilizan la minería de datos con el apoyo de la metodología CRISP DM, técnicas de segmentación y el software R con resultados satisfactorios.

Método

La investigación es de tipo cuantitativo no experimental ya que no se manipula deliberadamente variables, sino que se observan tal como se dan en su contexto natural para analizarlas, así mismo, es longitudinal de tipo panel porque se realizan tres mediciones a través del tiempo, una por el último semestre de 2022 y dos semestres por el año 2023 para analizar los cambios y evolución de las variables de los clientes activos y considera un algoritmo híbrido de segmentación utilizando métodos factoriales y el método de Ward, que según GAFILAT (2021, p. 52) es el más utilizado en las metodologías de lavado de dinero, asimismo, se apoya de la minería de datos utilizando la metodología CRISP-DM para la estructura y proceso de comprensión, preparación, modelado y evaluación de datos, organizada en:

Paso 1.- Comprensión del negocio:

La institución financiera tiene como finalidad contribuir al crecimiento económico impulsando a los sectores públicos y privados que generan divisas en el país. Estas actividades las lleva a cabo por medio del otorgamiento del crédito de forma directa o indirecta siendo éste un apoyo a través del otorgamiento de garantía sobre los créditos que la banca múltiple otorga al sector privado; así como, en el ámbito de mercados financieros a través de la mesa de dinero integrado por operaciones de reporto.

Paso 2.- Comprensión y preparación de los datos:

Con el propósito de asegurar su completitud, coherencia y precisión se realizan las siguientes actividades:

Recolección de datos

Los datos de las características inherentes de los clientes son recolectados por analistas y registrados en un sistema. La información que se captura en el sistema coincide con la información que se encuentra en el expediente de identificación integrado de manera física, ya que es validada por el área de prevención de lavado de dinero. Los datos transaccionales se obtienen por la extracción de información de los aplicativos transaccionales. A fin de comprobar la calidad de los datos, se considera la opinión favorable por parte del auditor externo de la revisión practicada para los ejercicios 2022 y 2023.

Selección de datos

Se consideran los factores de riesgo exigidos por las disposiciones en México que explican cómo y en qué medida cada cliente representa riesgo de lavado de dinero, siendo:

1. Tipo de persona: se considera la Evaluación Nacional de Riesgos 2020 (SHCP, 2020, p. 50) en la que menciona que la información requerida puede ser obstaculizada por la operación de empresas fachada, estructuras jurídicas complejas, fideicomisos y otros acuerdos legales que permiten una separación de la propiedad legal el uso de intermediarios ya que los esquemas o mecanismos más complejos podrían permitir operaciones con recursos de

- procedencia ilícita, se considera a las empresas y fideicomisos de riesgo alto y al resto de riesgo bajo.
2. Tipo de persona políticamente expuesta (PEP): aquella persona que desempeña o ha desempeñado funciones públicas destacadas en un país extranjero o en territorio mexicano conforme lo señalan las disposiciones y que en caso de ser extranjero se considera de riesgo alto.
 3. Antigüedad de la persona: la tipología emitida por la Unidad de Inteligencia Financiera considera a las personas jóvenes y entidades de reciente creación como el modo más común de operar para el lavado de dinero.
 4. Riesgo de la actividad económica.
 5. Ubicación geográfica: el crimen organizado y el narcotráfico suelen utilizar la economía para lavar dinero y sobornar a las autoridades incrementando así la tasa de criminalidad y la corrupción en distintas áreas geográficas.
 6. Producto utilizado: pueden aumentar el riesgo considerando: El tipo de persona con que se puede operar, tipo de moneda, las jurisdicciones involucradas en su operación, la complejidad del producto, la participación de intermediarios financieros en su operación y si la operación se pudiera dar de manera conjunta con otros bancos de desarrollo.
 7. Destino del recurso: indica si es un abono en moneda nacional (MN) o en dólares americanos (USD) relacionado con la actividad económica del cliente.
 8. Origen del recurso: indica si es nacional o extranjero relacionado con la actividad económica del cliente, ya que éstos pueden aumentar el riesgo considerando: La facilidad para la identificación de terceros involucrados, si permite o no la colocación o recepción de recursos en moneda extranjera, si permite o no la movilidad de recursos hacia o del extranjero y el tipo de proceso (presencial o no presencial que da origen a la operativa del canal).
 9. Número de operaciones: es la suma de operaciones realizadas en el semestre por parte del tipo de persona en el producto utilizado.
 10. Monto de operaciones.
 11. Número de contraparte: indica el número de cuentas bancarias de diferentes contrapartes a las que se depositarán los recursos, considerando de alto riesgo más de cinco cuentas.
 12. Transferencias de fondo: muestra el país al que se realiza la transferencia para analizar su perfil y comportamiento.

Limpieza de datos

Se lleva a cabo el análisis exploratorio para encontrar y eliminar registros incompletos, inexactos o incorrectos, “porque los métodos jerárquicos no tienen solución con valores perdidos y los valores atípicos deforman la distancia y producen clústeres unitarios” (Pérez, 2013, p. 279). A continuación, la Tabla 1 muestra el número de registros eliminados:

Tabla 1
Resumen general de los datos eliminados

Cientes con transacción	2022 2S	2023 1S	2023 2s
Total de registros	730	961	553
Total de registros eliminados	3	37	19
Porcentaje de representación de los registros eliminados	0.41%	3.85%	3.44%
Total de registros limpios	727	924	534

Nota. La tabla muestra el total de registros y la eliminación de registros por la inexactitud en el montón. El porcentaje de eliminación es bajo con respecto al total de los registros por lo que supone no afecta el resultado de la investigación.

Paso 3.-Ejecución:

Fase I.- Exploración de los datos

Con apoyo del software R se realiza el análisis descriptivo. Se identifica que el 53.95% son empresas y el 37.41% dependencias o entidades; solo uno es asimilado PEP extranjero; de las 324 actividades económicas en su mayoría son de riesgo bajo siendo: 20 % Sociedades de inversión, 9% instituciones financieras del extranjero y 7% servicio de la banca privada y mixta múltiple, entre otras; la antigüedad en promedio es de 22 años siendo su máximo 123; la ubicación geográfica se concentra en Ciudad de México y Nuevo León; el origen y destino del recurso principalmente es en moneda nacional y el 3.5% de las transferencias son a Estados Unidos. Se utilizan entre una a dos contrapartes para la dispersión de los recursos. En cuanto al número de operaciones, el mínimo es una y en promedio entre tres y cuatro. Respecto del monto de operaciones se puede observar la existencia de un sesgo positivo, ya que los montos registrados hasta el tercer cuartil son menores con respecto al máximo.

Fase II.- Segmentación de los factores clientes, productos, canales y ubicación geográfica: Técnica de Segmentación e identificación de factores

Se considera la clasificación sugerida por Pardo (2020) que, con base en lo revisado en las investigaciones precedentes, se robustece el método cuantitativo para esperar mejores resultados en la segmentación:

- Paso 1: se realiza el análisis en ejes principales correspondiente, utilizando por un lado el método de correspondencia simple para la segmentación de productos, canales y la segmentación integral de los cuatro factores; el objetivo del análisis de correspondencias simples es describir las asociaciones entre las variables fila y columna para tener una visión de los datos para su interpretación; por otro, el análisis factorial múltiple para la segmentación geográfica y clientes, considerando que la tabla de datos es multivariada. El análisis de segmentación es realizado a través de una clasificación a partir de las coordenadas de los primeros ejes obtenidos de los análisis factoriales. Estos segmentos se conforman de tal manera que los elementos dentro de cada segmento sean lo más parecido posible y que los elementos de diferentes segmentos sean lo más diferente posible. En el análisis es muy importante el concepto de la inercia o varianza total de los elementos a segmentar. Esta inercia se divide en inercia intraclases o intragrupos (varianza de los integrantes de un mismo segmento) e inercia interclases o intergrupos (varianza entre los centros de cada uno de los segmentos).
- Paso 2: se selecciona el número de ejes para la clasificación considerando varias opciones, por un lado, analizando las varianzas de las dimensiones principales con la función eig, los valores propios corresponden a la cantidad

de información retenida por cada dimensión; por otro lado, observando el gráfico de valores propios de las varianzas ordenadas de mayor a menor con la función `fviz_screeplot` del paquete de Factoextra en la que visualmente permite seleccionar las dimensiones observando la curvatura en la barra “llamada codo”; Para el caso de los factores productos y canales, se apoya adicionalmente con la calidad de representación de las filas y columnas del coseno al cuadrado; la representación gráfica de los punto de la fila y columna del coseno al cuadrado con el paquete Corrplot y la contribución de las filas y columnas a fin de explicar la variabilidad en el conjunto de datos. Para el caso de los factores clientes y geografía para confirmar el número de dimensiones se apoya del paquete Factoextra, la cual proporciona una lista de matrices que contiene todos los resultados de las variables activas (coordenadas, correlación entre variables y ejes, coseno al cuadrado y contribuciones) y con la calidad de representación de las variables en el mapa de factores coseno al cuadrado con el paquete Corrplot.

- Paso 3: se realiza la clasificación jerárquica con el método de Ward sobre los “individuos” del paso anterior. Este método necesita nuevamente la selección de un concepto de similitudes, disimilitudes o de distancias. Adicionalmente se requiere la selección de una distancia entre segmentos. Para determinar qué individuos se unen primero, se requiere calcular la matriz de distancias entre todas las parejas de individuos. Al unir la primera pareja se llega a una partición de $n-1$ (donde n es el número de individuos a segmentar) segmentos, una de ellas con dos individuos. Se requiere calcular la distancia entre el nuevo segmento y los individuos restantes. Al unir los dos individuos más próximos se tiene una nueva matriz de distancias de tamaño $(n-1, n-1)$. Sobre esta matriz se vuelve a seleccionar la pareja más próxima continuando el algoritmo así hasta llegar a un solo segmento que contiene todos los individuos.
- Paso 4: se decide el número de clases y se corta el árbol, para ello se utiliza la clasificación jerárquica. El dendrograma representa una serie de segmentos incorporados, en donde el número de segmentos decrece a medida que la altura del árbol crece. Para obtener una cantidad de segmentos particulares, se hace un corte en el árbol. Este árbol se construye partiendo del conjunto global de los individuos (clasificación descendente) y realizando divisiones sucesivas hasta llegar a cada uno de los individuos.
- Paso 5: se realiza un K-means de consolidación partiendo de los centros de gravedad de la partición obtenida al cortar el árbol.
- Paso 6: los segmentos obtenidos se caracterizan mediante la comparación de las estadísticas descriptivas al interior de ellos, con las estadísticas de la población clasificada. Para las variables continuas se compara la media al interior del segmento con la media general. Los perfiles de cada segmento vienen acompañados de la obtención de estadísticas descriptivas en su interior (principalmente la media y la desviación estándar) así como, de las estadísticas descriptivas de las variables ilustrativas de los individuos que pertenecen a cada segmento.

Para el factor clientes, los datos establecidos por las disposiciones para el riesgo inherente son: Tipo de persona, PEP, antigüedad del tipo de persona y riesgo de la actividad económica; por sus características transaccionales: Destino y origen del recurso, número y monto de operaciones realizadas en el semestre, número de contraparte y país en el que realiza transferencia de fondos.

Para el factor producto, se obtuvo la tabla binaria elaborada por el personal de desarrollo de productos en conjunto con el personal de prevención del lavado de dinero de la institución financiera validada por el auditor externo en la que indicaron con uno si el producto presenta el riesgo y cero de no presentarlo. El riesgo considerando: El tipo de persona con que se puede operar, tipo de moneda, las jurisdicciones involucradas en su operación, la complejidad del producto, la participación de intermediarios financieros en su operación y si la operación se pudiera dar de manera conjunta con otros bancos de desarrollo en México.

Para el factor canal, la institución financiera dispersa y recibe recursos financieros de sus clientes con cinco modalidades. Se obtuvo la tabla binaria realizada previamente por el personal de flujo de fondos en conjunto con el personal de prevención del lavado de dinero validada por el auditor externo en la que indicaron con uno si el canal presenta riesgo y cero en caso de no presentarlo. El riesgo considerando: La facilidad para la identificación de terceros involucrados, si permite o no la colocación o recepción de recursos en moneda extranjera, si permite o no la movilidad de recursos hacia o del extranjero y el tipo de proceso (presencial o no presencial que da origen a la operativa del canal).

La segmentación del factor de ubicación geográfica se realiza considerando todos los Estados que compone México, debido a que se realizan las transacciones desde cualquier estado. Los indicadores de riesgo utilizadas para la segmentación son: Homicidio siendo el número de víctimas de homicidio doloso por cada 100,000 personas; crimen violento siendo el número de delitos violentos por cada 100,000 personas, incluyen: Robo, asalto, violencia sexual y violencia dentro de la familia; y crimen organizado, que se compone de extorsiones, delitos mayores, delitos de drogas al por menor delitos y secuestro o trata de personas ;(IEP, 2023).

Los indicadores están en una escala de uno al cinco, en el que uno representa la puntuación más pacífica y cinco la menos pacífica. El crimen organizado y el narcotráfico suelen utilizar la economía para lavar dinero y sobornar a las autoridades incrementando así la tasa de criminalidad y la corrupción en consecuencia la posibilidad de ser sentenciados. Por otro lado, se considera el índice de competitividad estatal, la competitividad es la capacidad que tiene el estado para generar, atraer y retener talento e inversión, lo cual se traduce en productividad y bienestar para los habitantes e inversores.

Fase III.- Medición y Clasificación del Riesgo considerando un indicador de anormalidad:

El nivel de riesgo indica que tan crítico es el cliente considerando las diferentes amenazas implicadas en cada factor. Para medirlo, se elabora una matriz con el resultado de la segmentación, en el que se considera el valor de uno para riesgo bajo, dos para riesgo medio y el valor de tres para riesgo alto.

La Tabla 2 muestra los resultados en los que se observa que los factores: Clientes se encuentran en niveles de riesgo bajo y medio; Productos en su totalidad son ubicados como riesgo medio; Canales y ubicación geográfica presentan grados de riesgo alto y medio.

Tabla 2
Grados de Riesgo de los clientes por semestre

Factor	Periodo 2022 segundo semestre	Periodo 2023 primer semestre	Periodo 2023 segundo semestre
Cliente segmento 1	2	1	1
Cliente segmento 2	1	2	2
Cliente segmento 3	2	1	2
Cliente segmento 4	2	1	1
Producto segmento 1	2	2	2
Producto segmento 2	2	2	2
Producto segmento 3	2	2	2
Producto segmento 4	2	2	2
Canales segmento 1	2	2	2
Canales segmento 2	3	3	3
Canales segmento 3	2	2	2
U. Geográfica segmento 1	2	2	2
U. Geográfica segmento 2	3	3	3
U. Geográfica segmento 3	3	3	3
U. Geográfica segmento 4	2	2	2

La clasificación de riesgo es fundamental para asignar de forma eficaz los recursos disponibles de la institución financiera para mitigar el riesgo de LA/FT. Para clasificar al cliente por el grado de riesgo en cada semestre se realiza la suma del valor asignado en cada factor de riesgo, por lo que un nivel de riesgo alto su límite superior es 12 y el inferior 11; para riesgo medio alto su límite superior es 10 con un límite inferior nueve; riesgo medio su límite superior es ocho y límite inferior cinco; y para riesgo bajo es igual a 4. La Tabla 3 muestra el resumen de los resultados en los que se observa una mayor concentración de clientes en riesgo medio.

Tabla 3
Concentración de clientes en grados de riesgo por semestre

Periodo	Riesgo bajo	Riesgo medio	Riesgo medio alto	Riesgo alto
2022 segundo semestre	0	326	401	0
2023 primer semestre	0	850	74	0
2023 segundo semestre	0	437	97	0

La segmentación integra individuos homogéneos permitiendo identificar a los individuos atípicos dentro del segmento. Para analizar el comportamiento anormal es común utilizar en la gestión de riesgo dos desviaciones estándar del promedio del monto operado; por lo que se aplica dentro de la investigación considerando que es el empleado en el modelo actual de la institución financiera. Se lleva a cabo una segmentación integral de los cuatro factores creando grupos homogéneos en su interior y heterogéneos en su exterior para posteriormente generar una tabla con las operaciones de monto máximo, monto promedio y el monto promedio más dos desviaciones estándar por cliente en los tres periodos para comparar el monto de la operación promedio más dos desviaciones estándar y la operación máxima llevada por el cliente con la del segmento y así estar en condiciones de sospechar de una operación inusual.

Resultados

Para el factor cliente, el resultado son cinco segmentos para 2022, cuatro para el primer y segundo semestre de 2023, como ejemplo, para el segundo semestre de 2023 la figura 1 muestra el primer plano factorial en el que se observa la conformación de cuatro grupos asociados y la figura 2 muestra el dendrograma con cuatro segmentos conforme al análisis realizado junto con los valores propios/varianzas de las componentes principales, la gráfica de valores con la técnica del codo y el gráfico de correlación de variables. La Tabla 4 muestra el resultado de la clasificación jerárquica con el método de Ward y optimización de las clases con K-means.

Figura 1

Representación gráfica de los datos en dos dimensiones

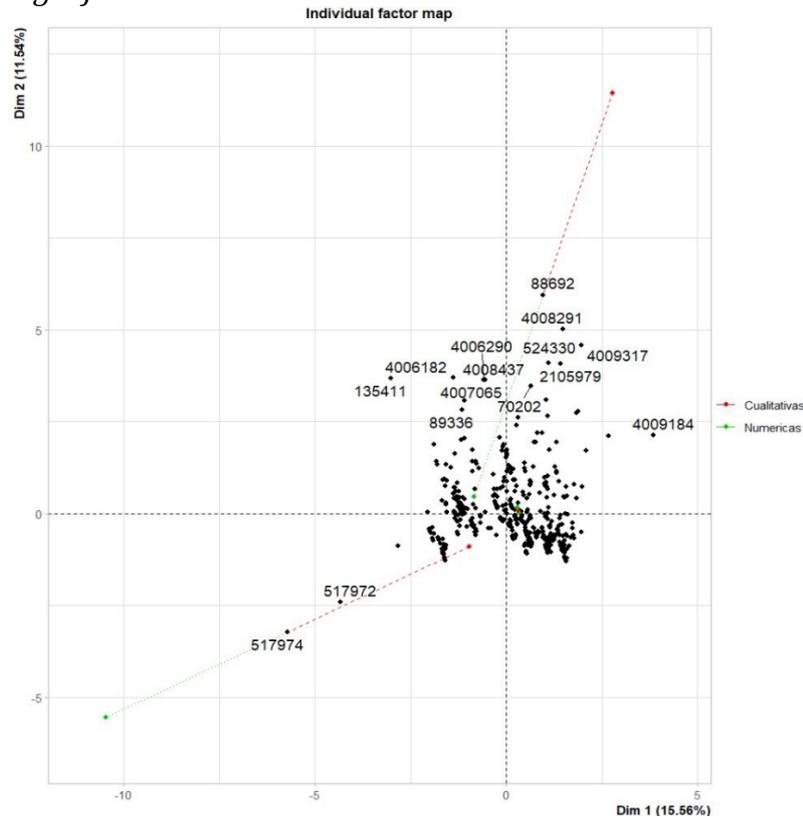


Figura 2
Representación gráfica del dendrograma

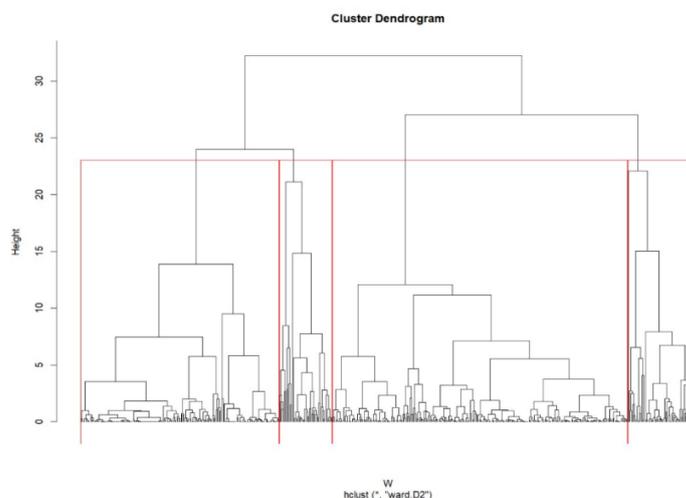


Tabla 4
Clasificación jerárquica con el método de Ward y optimización con K-means

Método	Segmento uno	Segmentos dos	Segmento tres	Segmento Cuatro
Ward	256	60	46	172
K-means	248	59	39	188

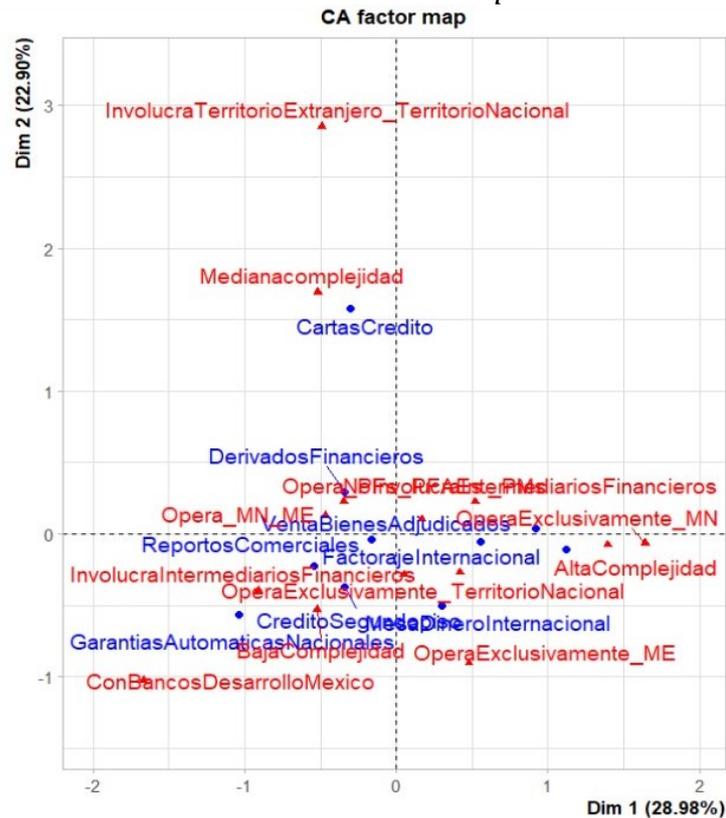
Los clientes que se agrupan en los diferentes segmentos en los tres semestres se parecen en su interior y se diferencian en su exterior, esto se corrobora con la estadística descriptiva que, por ejemplo, en el primer segmento agrupa el 46.44% del total de los clientes del segundo semestre de 2023 el 95.17% son empresas con el 57% del origen del recurso nacional y el 100% de su destino en moneda nacional; el riesgo de la actividad económica es medio; el número mínimo de operación es uno y su máximo 23; el promedio del monto de sus operaciones va a partir de \$1 hasta \$ 3,949,492,386 (MN) y el tercer segmento con el 35.21% de los clientes siendo el 91.44% dependencias y entidades con el 68% del origen del recurso extranjero y el 100% de su destino en moneda nacional; el riesgo de la actividad económica es baja; su número mínimo de operación es uno, pero su máximo es de 14 y el monto de sus operaciones van a partir de \$5,000 hasta \$286,160,000,000 (MN). Los clientes en el segundo segmento con características casi similares al primer segmento diferenciándose principalmente por el número y monto de la operación, sucede lo mismo para el tercer segmento con el cuarto segmento.

Para la segmentación de los productos, el plano factorial de la Figura 3 muestra que las cartas de crédito tiene riesgos diferentes respecto de los demás; así como, el de garantías automáticas nacionales y mesa de dinero internacional, se obtiene la calidad de las filas en la que resulta que nueve de los once productos están bien representados hasta la quinta dimensión al acercarse su suma a uno y solo dos: crédito de primer piso y reportos comerciales están lejanos; la gráfica de puntos de la fila en cinco dimensiones del coseno al cuadrado que muestra que sus valores mayores se encuentran dentro de las primeras cuatro dimensiones y la tabla de la contribución de las filas producto en porcentaje, que muestra los productos con mayor valor y que son las que más contribuyen a la definición de las dimensiones, siendo en la primera dimensión servicios fiduciarios y garantías automáticas nacionales y en la segunda dimensión cartas de crédito; por lo que respecta a la tabla de la calidad de las columnas la cual muestra si un riesgo está bien representado por las cinco dimensiones si la suma se acerca a uno, resultando que 12 de 14 están bien

representados y solo dos: opera en MN y ME; y opera exclusivamente ME están poco lejanos a uno y con la gráfica de los puntos de la columna en cinco dimensiones del coseno al cuadrado que muestra a todos los productos en el que se puede observar que sus valores mayores se encuentran dentro de las primeras tres dimensiones y el resultado por el algoritmo mixto son cinco segmentos obtenidos conforme con lo observado en el primer plano factorial y en el histograma junto con los valores propios/varianzas de las dimensiones principales y con la técnica del codo.

Figura 3

Primer plano factorial de los atributos de la variable producto



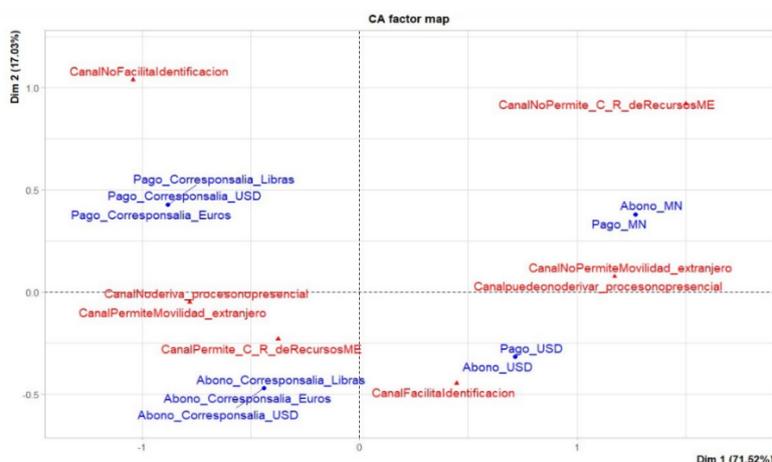
Con el resultado de la clasificación jerárquica con el método de Ward y optimización de las clases con K-means los productos contenidos en dos de los segmentos se parecen lo más posible diferenciándose de los tres productos que conformaron un segmento cada uno con base en el análisis realizado. Se evalúa la posibilidad de conformar cuatro segmentos en el que se observa que el producto de garantías automáticas nacionales forma parte del primer segmento afectando en la calidad de las filas.

Para la segmentación de los canales, el plano factorial de la Figura 4 muestra que los canales de pago de corresponsalía en libras, Euros y USD no facilitan la identificación, que los pagos y abonos en MN no permite la movilidad al extranjero y no son presenciales y que los abonos en corresponsalía en libras, Euros y USD y los pagos y abonos en USD facilitan la identificación de los clientes. Se obtiene la tabla del coseno al cuadrado de las filas canales, que muestra si un producto está bien representado por las tres dimensiones si la suma se acerca a uno, resultando que todos están bien representados, asimismo, la gráfica de los puntos de la fila en cinco dimensiones del coseno al cuadrado que muestra a todos los productos dentro de las tres dimensiones, la tabla de la contribución de las filas canales en porcentaje, que muestra los canales con mayor valor y que son las que más contribuyen a la definición de las dimensiones, siendo en la primera dimensión abono y

pago en MN y en la segunda dimensión los abonos en corresponsalía; resulta con las mismas técnicas la calidad de las columnas.

Figura 4

Primer plano factorial de los atributos de la variable canales

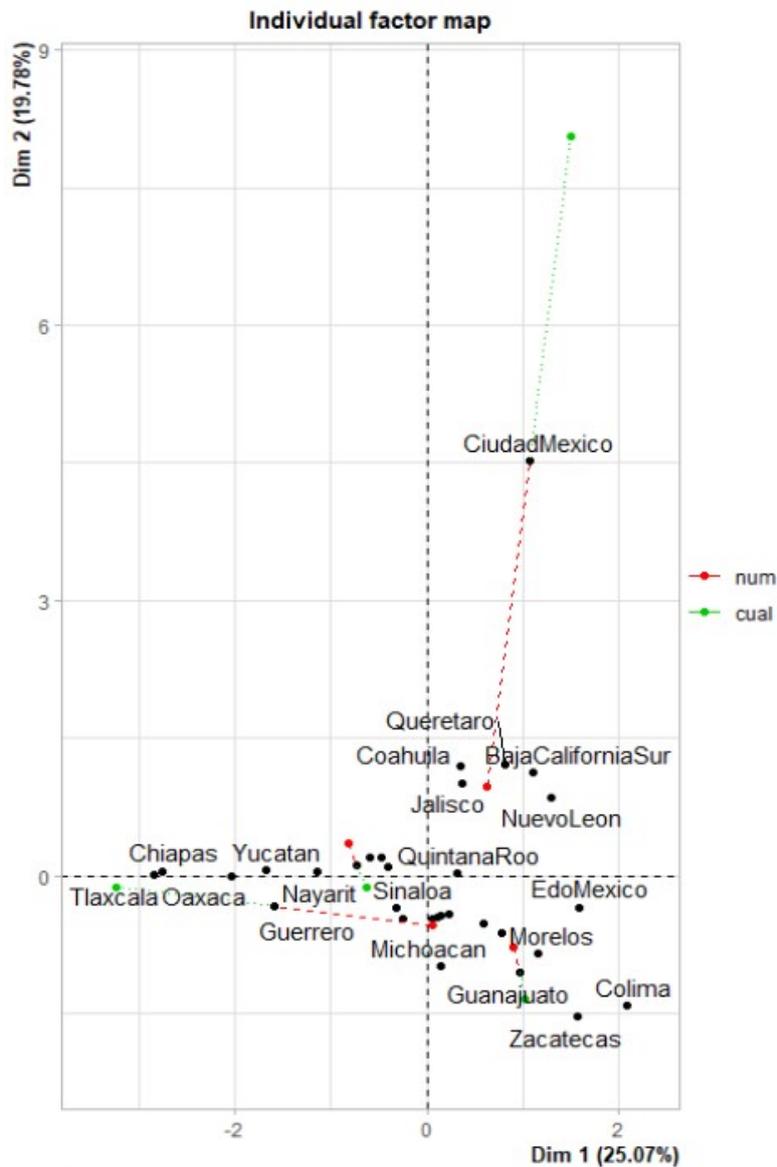


Con el método de Ward y optimización de las clases con K-means los canales contenidos en los tres segmentos se parecen lo más posible en su interior diferenciándose en su exterior con base en el análisis realizado. Se evalúa la posibilidad de conformar dos segmentos en el que se obtiene en el primer segmento a los pagos y abonos en MN y USD y en el segundo segmento los pagos y abonos en corresponsalía, este último se divide en dos en la propuesta; observando el diagrama simétrico del plano factorial el pago y abono en USD están más cercanos al abono en corresponsalía por lo que se supondría agrupaciones diferentes a las obtenidos, pero analizando la contribución y calidad de las filas y columnas en tres dimensiones se concluye que es la mejor opción.

La Figura 5 muestra el plano factorial de la ubicación geográfica que permite observar la conformación de cuatro grupos, los cuales se corroboraron con la obtención de los valores propios/varianzas, la gráfica de valores ordenadas de mayor a menor y la gráfica de puntos en las dimensiones del coseno al cuadrado que muestra que los mayores valores de las variables numéricas se encuentran en la primera y hasta la cuarta dimensión para la variable cualitativa y con el resultado de la clasificación jerárquica con el método de Ward y optimización de las clases con K-means las ubicaciones geográficas contenidas en los cuatro segmentos se parecen lo más posible en su interior diferenciándose en su exterior con base en el análisis realizado. Los segmentos de mayor riesgo son el segundo integrado por los estados de Baja California Sur, Coahuila, Jalisco, Nuevo León y Querétaro, los cuales se caracterizan por tener un índice de competitividad Alto; y el tercer segmento compuesto por los estados de Campeche, Colima, Guanajuato, Hidalgo, Michoacán, Morelos, Puebla, San Luis Potosí, Estado de México, Tabasco, Veracruz y Zacatecas. Se caracterizan por tener un índice de competitividad medio bajo y esto supone que es ocasionado por ser de los estados con mayor percepción en delincuencia e inseguridad; integra a la Ciudad de México con un índice de competitividad muy alto.

Figura 5

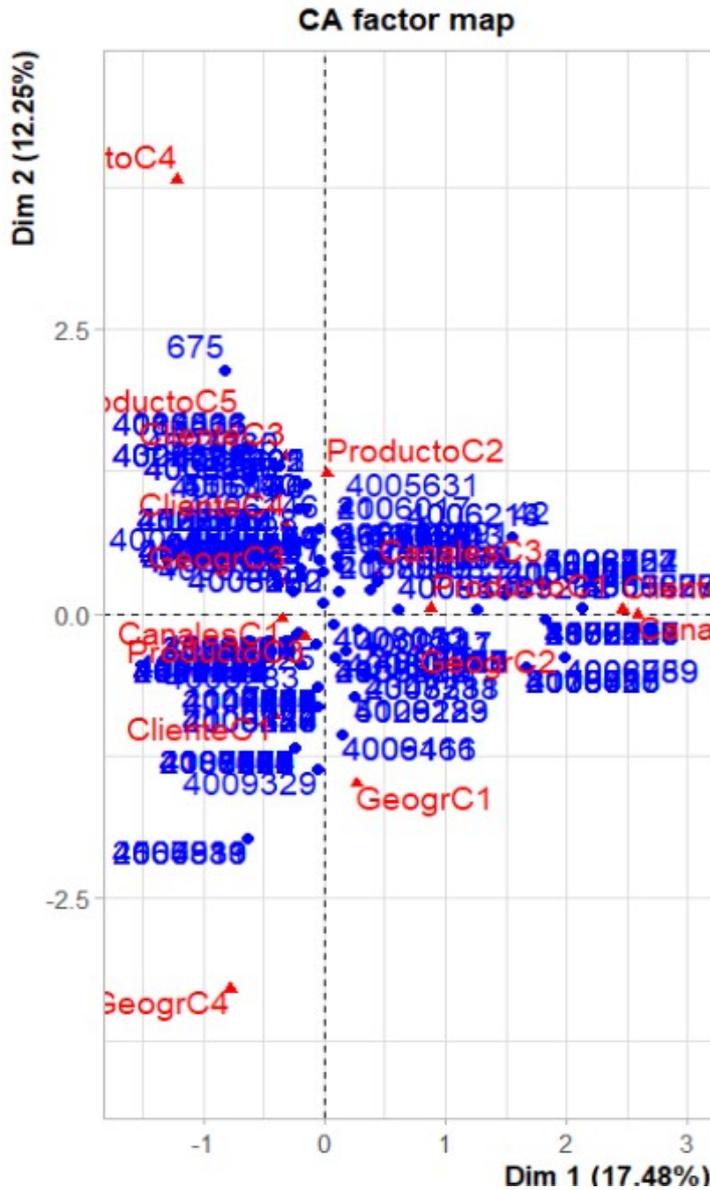
Primer plano factorial de los atributos de la variable ubicación geográfica



Para concluir se realiza la integración de la segmentación de los cuatro factores por cada semestre aplicando un análisis de correspondencias, la Figura 6 se muestra el plano factorial obtenido en el que se observan la concentración de la nube de puntos en cuatro grupos, los puntos y números de color azul representan las clientes, mientras que los de color rojo la ubicación de los factores de riesgo. Estos resultados fueron validados con el histograma junto con los valores propios/varianzas de las dimensiones principales y su gráfica con la técnica del codo y con el resultado de la clasificación jerárquica con el método de Ward y optimización de las clases con K-means.

Figura 6

Primer plano factorial de los cuatro segmentos del segundo semestre de 2023



La Tabla 5 muestra los resultados de los clientes que presentaron alerta y que considerando los montos agrupados de la segmentación se establece un umbral del monto máximo de operaciones en comparación con los del cliente en ese periodo y periodos anteriores lo que permite conocer la transaccionalidad del cliente y aceptar o rechazar la alerta y así reducir los falsos positivos.

Tabla 5
Cientes con alerta

Periodo	Total	Con alerta	Confirma la alerta	Alerta riesgo medio	Alerta riesgo medio alto
2022 segundo semestre	727	1	0	0	1
2023 primer semestre	924	518	14	13	1
2023 segundo semestre	534	183	15	6	9

Discusión y conclusiones

En conclusión, México en relación con el LA/FT, de acuerdo con BASEL, enfrenta una situación de riesgo medio y, por lo tanto, requiere fortalecer sus estrategias de prevención, control y detección.

Para las organizaciones delictivas, las instituciones financieras son de utilidad e interés, ya que su conexión internacional les permite operar grandes cantidades de dinero por todo el mundo y así facilitarles su acción, es por ello la importancia y responsabilidad de las instituciones financieras de actuar de forma eficaz y oportuna en su gestión preventiva (Guevara y Flores, 2021, p. 5).

Se cumple con el objetivo y preguntas de investigación planteadas ya que la técnica de segmentación sugerida por Pardo (2020) permite agrupar, explicar y validar la integración homogénea en su interior y heterogénea en su exterior, así como medir y explicar el grado de riesgo final del cliente. Con los datos recopilados durante los tres periodos en conjunto con los valores agrupados en cada segmento permite establecer un umbral y comprar la transaccionalidad de los clientes para tener mayor conocimiento, comparabilidad y elementos para determinar la anormalidad de las operaciones y con ello decidir de una mejor manera las operaciones inusuales con un enfoque basado en riesgo y realizar una gestión de conocimiento más específica de los clientes.

Los resultados del método de segmentación propuesto reflejan un perfil de riesgo medio y alto de los clientes de la institución financiera en materia de lavado de dinero contrario al método actualmente utilizado en el que concentra en riesgo bajo. Esto es explicado por la posición geográfica de México y los riesgos derivados de la colindancia con otras jurisdicciones (Secretaría de Hacienda y Crédito Público, 2020) y los indicadores tomados de IEP (2023), así mismo, por los riesgos agrupados en los factores de productos y canales utilizados junto con las características inherentes y transaccionales de los clientes.

El presente artículo propone un modelo que considera las recomendaciones descritas en los informes del GAFI, mejores prácticas y las disposiciones nacionales; así como variables que no se habían incluido en estudios previos, la asignación de un grado de riesgo y umbrales para la identificación de operaciones sospechosas y por lo tanto mejorándolo.

Una de las limitaciones del modelo es la dependencia de la calidad de datos, por lo que se recomienda llevar a cabo el proceso de limpieza y; establecer y mantener adecuadamente la calidad de los datos para estar en condiciones de impedir el buen funcionamiento del Método, asimismo, es la ausencia de datos de clientes confirmados en lavado de dinero para conocer y validar el comportamiento del modelo.

Conviene recordar las directrices del BASEL (2020) en la que es fundamental que las instituciones financieras comprendan y documenten la actividad bancaria normal y la esperada de sus clientes. Esto permitirá establecer mecanismos efectivos para detectar las operaciones que se desvíen del patrón habitual de actividad bancaria.

Para investigaciones futuras, se sugiere incorporar indicadores de la actividad económica del país que permita compararlo con el comportamiento del cliente y del segmento y valorar si los incrementos o disminuciones son coherentes entre ellos para una mejora en el Método.

Referencias

- Basel Institute on Governance. (2020). *Sound management of risks related to money laundering and financing of terrorism: revisions to supervisory cooperation [Gestión adecuada de los riesgos relacionados con el blanqueo de capitales y la financiación del terrorismo: revisiones de la cooperación supervisora]*. <https://www.bis.org/bcbs/publ/d505.htm>
- Basel Institute on Governance (BASEL). (2022). *Basel AML Index 2022: 11th Public Edition - Ranking money laundering and terrorist financing risks around the world*. <https://baselgovernance.org/publications/basel-aml-index-2022>
- Basel Institute on Governance (BASEL). (2023). *Basel AML Index 2023: 12th Public Edition - Ranking money laundering and terrorist financing risks around the world*. <https://baselgovernance.org/publications/basel-aml-index-2023>
- Camacho, M., Rocha, J., & Segovia, M. (2021). Money laundering and terrorism financing detection using neural networks and an abnormality indicator. *Expert Systems with Applications*, 169, 114470. <https://doi.org/10.1016/j.eswa.2020.114470>
- Castro, N., & Castro, M. (2020). *Segmentación de clientes en un fondo de empleados para identificar los riesgos de lavado de activos y financiación del terrorismo*. <https://repository.libertadores.edu.co/items/04da680f-e224-4b18-a09f-9cce5ebc0706>
- Comisión Nacional Bancaria y de Valores (CNBV). (2020). *Guía para el Fortalecimiento del Régimen Preventivo de Lavado de Dinero y Financiamiento al Terrorismo, derivada de los Hallazgos recurrentes detectados durante los procesos de Supervisión: Apartado 3. Conocimiento del Cliente y Debida Diligencia Reforzada*. https://issuu.com/gmc360/docs/2020_07_15_gui_a_hallazgos_recurrentes
- Grupo de Acción Financiera de Latinoamérica (GAFILAT). (2021). *Análisis estratégico sobre el uso de Inteligencia Artificial, Minería de Datos y Análisis de Big Data en prevención y detección LA/FT (UIF/MP)*. <https://biblioteca.gafilat.org/?cat=17>
- Grupo de Acción Financiera de Latinoamérica (GAFILAT). (2023). <https://www.gafilat.org/index.php/es/inicio>
- Guevara Rabanal, M. L., y Flores Tananta, D. C. A. (2021). Lavado de activos y su relación con la rentabilidad, banco de la nación, 2020. *Ciencia Latina Revista Científica Multidisciplinar*, 5(5), 8212-8227. https://doi.org/10.37811/cl_rcm.v5i5.899
- Institute for Economics & Peace (IEP). (2023). *Mexico Peace Index 2023: Identifying and Measuring the Factors That Drive Peace*, Sydney, May 2023. <http://visionofhumanity.org/resources>
- Jovel, T. (2020). *Desarrollo de un modelo analítico para la segmentación de asociados en una cooperativa de ahorros y crédito*. [Tesis Maestría, Universidad Nacional de Colombia]. <https://repositorio.unal.edu.co/handle/unal/79648>
- Martínez, G., Álvarez, Y. I., y Silva, H. (2022). Empresas y prevención del lavado de dinero en México. *3C Empresa. Investigación y pensamiento crítico*, 11(1), 67-83. <https://doi.org/10.17993/3cemp.2022.110149.67-83>
- Pardo, C. (2020). *Estadística descriptiva multivariada*. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/79914>
- Pérez, L. (2013). *Técnicas de Segmentación. Conceptos, herramientas y aplicaciones*. Alfaomega Grupo Editor, S.A. de C.V
- Perez, L. (2020). *Metodología para segmentación de un SARLAFT*. https://repositorio.unbosque.edu.co/bitstream/handle/20.500.12495/6168/Perez_Lincoln_Ernesto_2020.pdf?sequence=2&isAllowed=y

Secretaría de Hacienda y Crédito Público. (2020). *Evaluación Nacional de Riesgos. Versión Pública del mes de septiembre de 2020.*
<https://www.pld.hacienda.gob.mx/work/models/PLD/documentos/enr2020.pdf>

Secretaría de Hacienda y Crédito Público (SHCP). (2022). *Disposiciones de Carácter General a que se refiere el Artículo 115 de la Ley de Instituciones de Crédito. Última Reforma publicada en el Diario Oficial de la Federación el 03 de marzo de 2022.*
https://www.gob.mx/cms/uploads/attachment/file/709513/DCG_Compiladas_Instituciones_de_Credito_08.03.2022.pdf