Clasificación de perfiles de riesgo de usuarios basados en su comportamiento transaccional para identificar posibles casos relacionados con el lavado de activos y/o financiación del terrorismo mediante algoritmos no supervisados en Nequi, periodo octubre 2023 - mayo 2024.

Miguel Fernando Sosa Zapata
Stiven Cadavid Cataño

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor
Javier Fernando Botia Valderrama, Doctor en Ingeniería Electrónica

Universidad de Antioquia
Facultad de Ingeniería
Especialización en Analítica y Ciencia de Datos
Medellín, Antioquia, Colombia
2024

Cita	(Sosa Zapata & Cadavid Cataño, 2024)
Referencia **Estilo APA 7 (2020)**	Sosa Zapata, M., & Cadavid Cataño, S. (2024). *Clasificación de perfiles de riesgo de usuarios basados en su comportamiento transaccional para identificar posibles casos relacionados con el lavado de activos y/o financiación del terrorismo mediante algoritmos no supervisados en Nequi, periodo octubre 2023 - mayo 2024.* [Trabajo de grado especialización]. Universidad de Antioquia, Medellín, Colombia.

Especialización en Analítica y Ciencia de Datos, Cohorte VII.

Centro de Investigación Ambientales y de Ingeniería (CIA).

Dedicatoria

A nuestras familias que son el pilar de todo.

Agradecimientos

A los tutores y compañeros que han contribuido a nuestra formación.

Tabla de contenido

Contenido

Lista de tablas

Lista de figuras

Siglas, acrónimos y abreviaturas

LA/FT	Lavado de Activos / Financiación del Terrorismo
PIB	Producto Interno Bruto
SARLAFT	Sistema de Administración de Riesgo de LA/FT
UIAF	Unidad de Información y Análisis Financiera

Resumen

Este estudio se enfoca en la identificación de comportamientos transaccionales anómalos en los usuarios de Nequi, con el fin de detectar posibles señales de lavado de activos y financiación del terrorismo. En un entorno de creciente digitalización y complejidad en el sector financiero, es esencial contar con métodos eficaces para monitorear y clasificar las transacciones sospechosas. Para abordar este desafío, se utilizó un conjunto de datos de 582.710 registros, del cual se generaron conjuntos de entrenamiento y validación mediante la selección de IDs únicos, optimizando así la capacidad de generalización del modelo. Las métricas de desempeño utilizadas incluyen el Índice de Silueta, homogeneidad, completitud y distancia de separación, con el fin de evaluar la efectividad de los algoritmos para distinguir entre instancias anómalas y no anómalas. El modelo se ajustó de manera iterativa utilizando Isolation Forest y Fuzzy C-means, logrando un incremento en el Índice de Silueta de al menos 0.8, lo que asegura una clasificación precisa de los perfiles de riesgo y reduce el tiempo necesario para identificar actividades ilícitas.

Posterior a la detección de anomalías mediante algoritmos de agrupamiento, se entrenaron algoritmos de clasificación para estimar los niveles de riesgo de los usuarios desde un enfoque probabilístico, clasificándolos en categorías que permiten priorizar las investigaciones según el perfil de riesgo. Esto contribuye a una asignación eficiente de recursos en el área de cumplimiento, optimizando el tiempo y los esfuerzos dedicados a la prevención de actividades ilícitas.

*Palabras clave***:** Sarlaft, aprendizaje no supervisado, entidad bancaria, riesgos financieros, transacciones.

Abstract

This study focuses on identifying anomalous transactional behaviors in Nequi users to detect potential signs of money laundering and terrorist financing. In an environment of increasing digitalization and complexity in the financial sector, it is essential to have effective methods to monitor and classify suspicious transactions. To address this challenge, a dataset of 582,710 records was used, from which training and validation sets were created by selecting unique IDs, optimizing the model's generalization capability. The performance metrics used include the Silhouette Index, homogeneity, completeness, and separation distance, to assess the effectiveness of the algorithms in distinguishing between anomalous and non-anomalous instances. The model was iteratively tuned using Isolation Forest and Fuzzy C-means, achieving an increase in the Silhouette Index of at least 0.8, ensuring accurate risk profile classification and reducing the time required to detect illicit activities.

After anomaly detection through clustering algorithms, classification algorithms were trained to estimate users' risk levels using a probabilistic approach, categorizing them into risk profiles that allow prioritization of investigations based on the risk level. This contributes to an efficient allocation of resources in the compliance department, optimizing time and efforts dedicated to preventing illicit activities.

*Keywords***:** money-laundering, unsupervised learning, banking, financial risks, transactions.

1. Descripción del problema

1.1. Problema de negocio

La lucha contra el lavado de activos es una prioridad global debido a su enorme impacto en la economía y la sociedad. Según la Unidad de Información y Análisis Financiera (UIAF, 2014), se estima que las transacciones de lavado de activos representan entre el 2% y el 5% del Producto Interno Bruto (PIB) mundial. Esta cifra alarmante subraya la urgencia de implementar sistemas efectivos de prevención y detección en instituciones financieras.

Para empresas financieras como Nequi Compañía de Financiamiento S.A., el desafío principal radica en clasificar y monitorear las transacciones de sus usuarios mediante perfiles de riesgo. Utilizando bases de datos transaccionales recientes, Nequi puede identificar patrones sospechosos de manera temprana y actuar con celeridad ante posibles casos de lavado de activos, mitigando posibles riesgos financieros y reputacionales.

La investigación de Masciandaro (2013) sobre el "efecto estigma" amplía la comprensión de los riesgos asociados con el incumplimiento de las normativas internacionales contra el lavado de dinero. Su estudio revela que los países incluidos en "listas negras" de jurisdicciones no cooperativas experimentan una disminución en el flujo de capital, un incremento en los costos de cumplimiento y una considerable carga reputacional para sus sistemas financieros. Esta estigmatización funciona como un recordatorio poderoso de la importancia de la transparencia y el cumplimiento normativo, no sólo para evitar sanciones, sino también para fortalecer la confianza en el sistema financiero. En este sentido, empresas como **Nequi** desempeñan un papel clave al implementar sistemas robustos de monitoreo transaccional y cumplimiento.

Según Buchanan (2004), la globalización ha facilitado la comunicación entre los lavadores de dinero, permitiéndoles distribuir sus operaciones en un mayor número de jurisdicciones. Este fenómeno incrementa el número de obstáculos legales para las investigaciones, complicando la identificación y el rastreo de actividades sospechosas a nivel internacional.

Entre 2012 y 2013, varios bancos internacionales, incluyendo Royal Bank of Scotland, Standard Chartered, Unicredit Group, Barclays, Hong Kong Shanghai Banking Corporation (HSBC), JPMorgan Chase y Citigroup, fueron investigados, acusados, multados y/o obligados a mejorar el cumplimiento de regulaciones en relación con transacciones financieras ilícitas (Powell, 2013). Estos episodios ilustran el riesgo significativo que implica la falta de cumplimiento y la necesidad de adoptar mecanismos de prevención y monitoreo adecuados.

Para Nequi, el objetivo es utilizar estos conocimientos para desarrollar modelos de clasificación de riesgo precisos y eficientes que permitan detectar de manera proactiva transacciones relacionadas con el lavado de dinero y la financiación del terrorismo. La medición de la precisión y eficiencia en la identificación temprana de transacciones fraudulentas será crucial en este proceso. Con una estrategia bien fundamentada en el análisis de datos y alineada con las mejores prácticas internacionales, Nequi puede no solo protegerse del riesgo de lavado de activos, sino también contribuir a la estabilidad y seguridad del sistema financiero global.

1.2. Aproximación desde la analítica de datos

El problema se abordará aplicando algoritmos de aprendizaje no supervisado para segmentar a los clientes según su comportamiento transaccional a lo largo del tiempo. Se seleccionará el modelo de agrupamiento más adecuado para las características específicas de los datos.

Huang, *et al* (2024) proponen el uso del K-means basado en aprendizaje automático para mejorar la detección de fraudes financieros. Los resultados experimentales confirman que K-means es especialmente útil en áreas de alto riesgo, proporcionando a las instituciones financieras herramientas más precisas para monitorear y prevenir actividades fraudulentas.

Por su parte, en la banca han usado diferentes metodologías que combinan características de nodos obtenidas de egonets [1] reducidas, es decir, subgrafos que excluyen conexiones menos relevantes, y emplea el algoritmo Isolation para identificar nodos (cuentas) con comportamientos

[1] **egonet** es el subgrafo de un nodo en el que se incluyen tanto el nodo en cuestión como todos sus vecinos directos (tanto los nodos de entrada como los de salida). Este subgrafo incluye todos los nodos que están directamente conectados al nodo central.

anómalos Liu *et al*. (2008). Los resultados de la investigación muestran que este enfoque es eficaz y robusto, superando a otros métodos en la detección de patrones sospechosos tanto en datos reales como en datos sintéticos de transacciones. Este método se presenta como un complemento adecuado a los sistemas basados en reglas, ya que permite identificar esquemas de fraude complejos que los enfoques tradicionales suelen pasar por alto (Dumitrescu, *et al*, 2022).

Por otro lado, Chen, *et al* (2011) destacan que el modelo de inferencia difusa basado en reglas permite evaluar transacciones utilizando reglas difusas diseñadas con el conocimiento de un humano experto de patrones en lavado de dinero. Estas reglas clasifican las actividades transaccionales en función de la cantidad de dinero y el tiempo entre las transacciones. Al definir grados de sospecha para distintas combinaciones de variables, el sistema asigna un puntaje de riesgo a cada transacción, facilitando la identificación de posibles actividades de lavado de dinero.

En resumen, nuestros hallazgos pueden ayudar a definir políticas de monitoreo para desarrollar acciones preventivas y reducir el lavado de dinero en corresponsales no bancarios. Sin embargo, es importante contar con el apoyo constante de las instituciones financieras y de personal experto para contrarrestar una situación en crecimiento.

1.3. Origen de los datos

Los datos provienen de muestras transaccionales capturadas por la aplicación. Estas muestras incluyen interacciones entre clientes, compras, identificadores y montos transados. La fuente de datos se encuentra almacenada en buckets de S3, en formato *parquet* particionado por año, mes y día.

1.4. Métricas de desempeño

Las métricas de desempeño incluirán, desde el punto de vista del aprendizaje automatico, el Índice de Silueta, cuyo valor cercano a 1 indicará que los grupos están bien agrupados y claramente separados entre sí. Adicionalmente, se utilizarán métricas de homogeneidad y completitud para evaluar la consistencia y cohesión de los grupos, así como la distancia de separación entre instancias, una medida clave en algoritmos de detección de anomalías para asegurar una adecuada diferenciación entre comportamientos normales y anómalos.

En cuanto a los modelos de clasificación, el objetivo será optimizar métricas como el puntaje F1, para equilibrar precisión y exhaustividad, y el AUC, que permite medir la capacidad del modelo para distinguir entre casos positivos y negativos de manera robusta. Asimismo, se explorará el ajuste de otros parámetros, como el valor predictivo negativo, para minimizar falsos positivos en contextos de alta criticidad.

Finalmente, desde la perspectiva del negocio, se medirá el impacto del modelo en la reducción del tiempo de detección de fraudes, una métrica crucial para mejorar la eficiencia operativa y la capacidad de respuesta ante actividades ilícitas.

2. Objetivos

2.1. Objetivo general

Proponer una metodología que permita identificar los usuarios de Nequi en perfiles de riesgo basados en su transaccionalidad, los posibles lavados de activos y la sospecha de financiación al terrorismo para un periodo de observación de 10 meses.

2.2. Objetivos específicos

- Analizar la base de datos de los usuarios de Nequi durante un periodo de observación de 10 meses mediante una exploración de datos y el conocimiento previo del escenario de aplicación.
- Evaluar los modelos de agrupamiento de datos mediante el índice de la silueta a partir de un umbral de decisión de 0.8
- Encontrar el mejor modelo de clasificación mediante el uso de métricas de validación y un umbral de decisión del puntaje F1 mayor o igual a 0.8.

3. Datos

3.1. Datos originales

A continuación, se aclara que los datos utilizados en este análisis son **confidenciales**. Su uso fue autorizado exclusivamente para este ejercicio académico bajo estrictas restricciones, **prohibiendo su difusión o intercambio con terceros** ajenos al proyecto.

Los datos crudos se encuentran almacenados en Buckets del servicio S3 de AWS en formato Parquet de forma particionada y ocupan aproximadamente 36.9 megabytes. La información transaccional, almacenada en una zona curada, se deriva de las actividades de 993 individuos durante los últimos 10 meses. Los datos están particionados por año, mes y día, lo que facilita su acceso y análisis. Cada registro incluye columnas como el ID del individuo, fecha de la transacción, monto de la transacción, categoría de la transacción y método de pago. Además, los datos permiten el uso de algoritmos de aprendizaje no supervisado, proporcionando una base sólida para análisis y modelos predictivos avanzados. Finalmente, se consolidó un dataset con una muestra de 582,710 registros y 57 variables; se comparte en el anexo 1 el diccionario de los campos útiles desde la muestra a trabajar. Para este caso, los datos se encuentran anonimizados dada las restricciones legales y las directrices internas establecidas desde el Gobierno de Datos de la compañía respecto al tratamiento de información sensible.

3.2. Bases de datos

A partir de los datos crudos disponibles, se realiza un procesamiento que incluye agrupaciones y la creación de variables dummies para atributos como el día y la hora de la transacción. El objetivo es obtener un registro único de cada cliente que consolide toda su actividad transaccional: cantidad de transacciones por hora del día, cantidad de transacciones por día de la semana, transacciones de débito y crédito, montos transados, número de cuentas con las que el usuario ha transaccionado en el periodo analizado, entre otras características, de acuerdo con lo mencionado por Chen et al. (2018), donde expresa la necesidad de delimitar un horizonte de tiempo para tener una vista holística de todas las actividades transaccionales y apoya el hecho de

consolidar las transacciones crudas por individuo como el monto total enviado en un período, para construir perfiles de clientes.

Posteriormente, se construyen las bases de datos de entrenamiento, prueba y validación para optimizar la capacidad de generalización del modelo. Para lograr esto, se realiza una división de los datos en la que los identificadores de los individuos en el conjunto de entrenamiento son exclusivos respecto a los de los conjuntos de prueba y validación. Así, el conjunto de entrenamiento contiene el 70% de los datos, mientras que el conjunto de prueba y el conjunto de validación abarcan el 20% y el 10%, respectivamente.

Para esta división, se selecciona aleatoriamente un subconjunto de identificadores únicos para cada conjunto, asegurando que no haya solapamiento entre los individuos. Este enfoque permite evaluar con mayor precisión la capacidad del modelo para generalizar sobre datos nuevos, maximizando su efectividad en escenarios no observados.

3.3. Analítica descriptiva

La siguiente tabla permite visualizar el conjunto de datos que permitirá cumplir con el objetivo de establecer un análisis detallado de las transacciones realizadas, identificando patrones, comportamientos y posibles anomalías en las transacciones financieras.

Tabla 1

Previsualización del dataset definitivo

dia_transaccion	codigo_producto	hora_transaccion	valor_transaccion	naturaleza_transaccion	codigo_concepto	concepto_desc	canal_desc	codigo_desc	entidad_contraparte
2024-02-08	SHCAS	09:31:59	4000.000	CREDITO	T001	Transferencia entre cuentas Nequi	NEQUI TRANSFERENCIA		Bancolombia Banca Digital
2024-02-08	SHCAS	11:26:30	9000.000	DEBITO	T001	Transferencia entre cuentas Nequi	NEQUI TRANSFERENCIA		Bancolombia Banca Digital
2024-02-08	SHCAS	16:27:51	50000.000	DEBITO	C005				
2024-02-08	SHCAS	13:49:42	30000.000	DEBITO	T001	Transferencia entre cuentas Nequi	NEQUI TRANSFERENCIA		Bancolombia Banca Digital
2024-02-08	SHCAS	12:03:01	256500.000	DEBITO	T001	Transferencia entre cuentas Nequi	NEQUI TRANSFERENCIA		Bancolombia Banca Digital

El análisis de las transacciones a diferentes contrapartidas se ilustra a través de dos gráficos: un diagrama de caja y bigotes y un histograma. Estos gráficos nos brindan información valiosa sobre la distribución de las transacciones *véase figuras 1 y 2*.

Los puntos situados fuera de los bigotes del boxplot representan valores atípicos. En este caso, hay numerosos valores atípicos que se extienden hasta 4232 transacciones, una mediana transaccional de 36 y una media de 92. Esto sugiere que hay algunas contrapartidas con un número excepcionalmente alto de transacciones, mientras que la mayoría de las contrapartidas tienen un número mucho menor de transacciones.

Figura 1

Distribución del conteo de transacciones - Diagrama de cajas

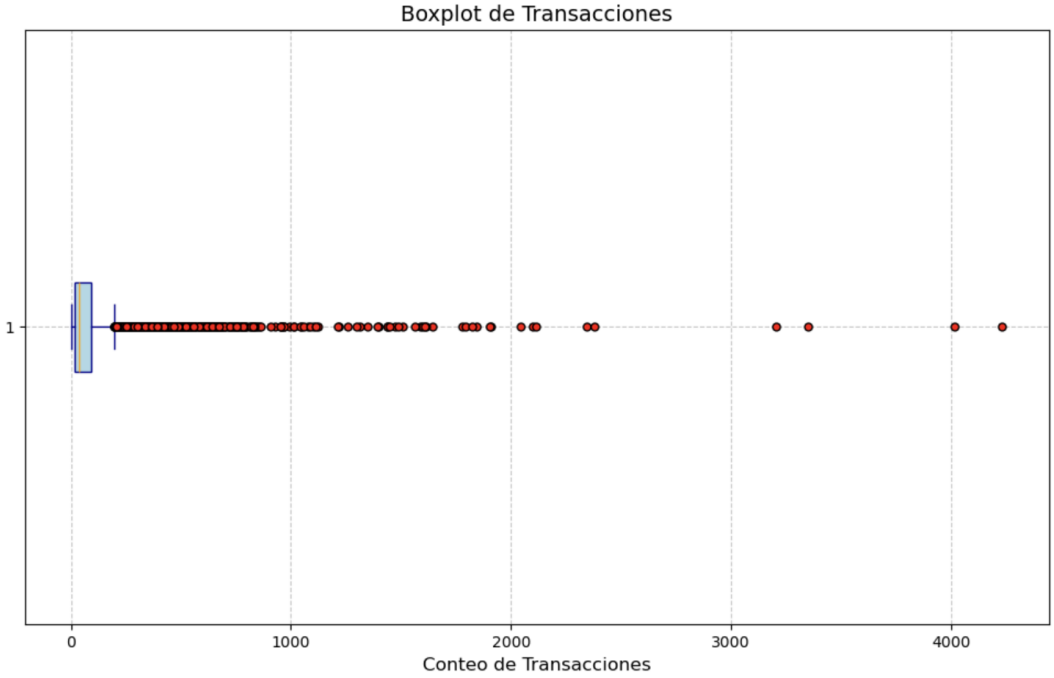
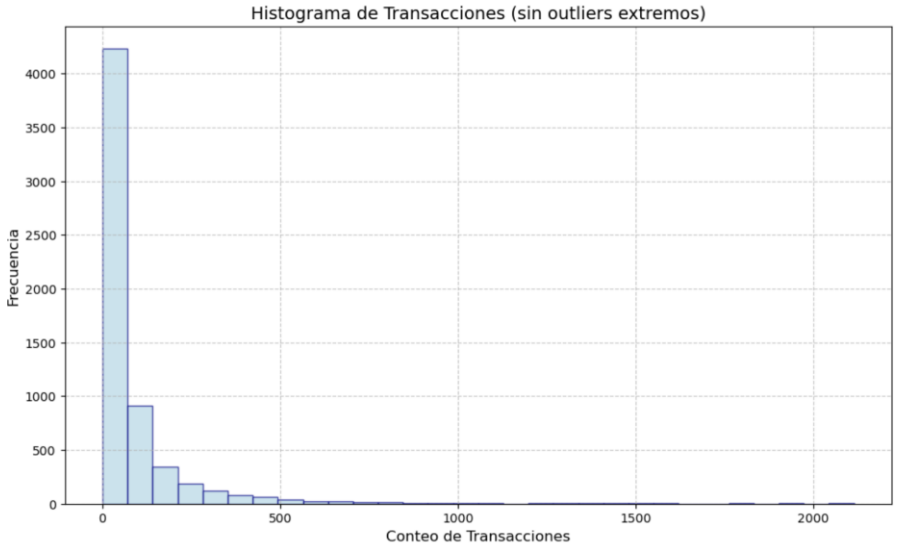


Figura 2

Distribución del conteo de transacciones - Histograma



Analizando el valor transaccional habitual en la muestra tomada, se observan una mediana de 17.000 y una distribución muy sesgada a la derecha a causa de datos atípicos que se observan a continuación.

Figura 3

Distribución del valor de las transacciones - Diagrama de cajas

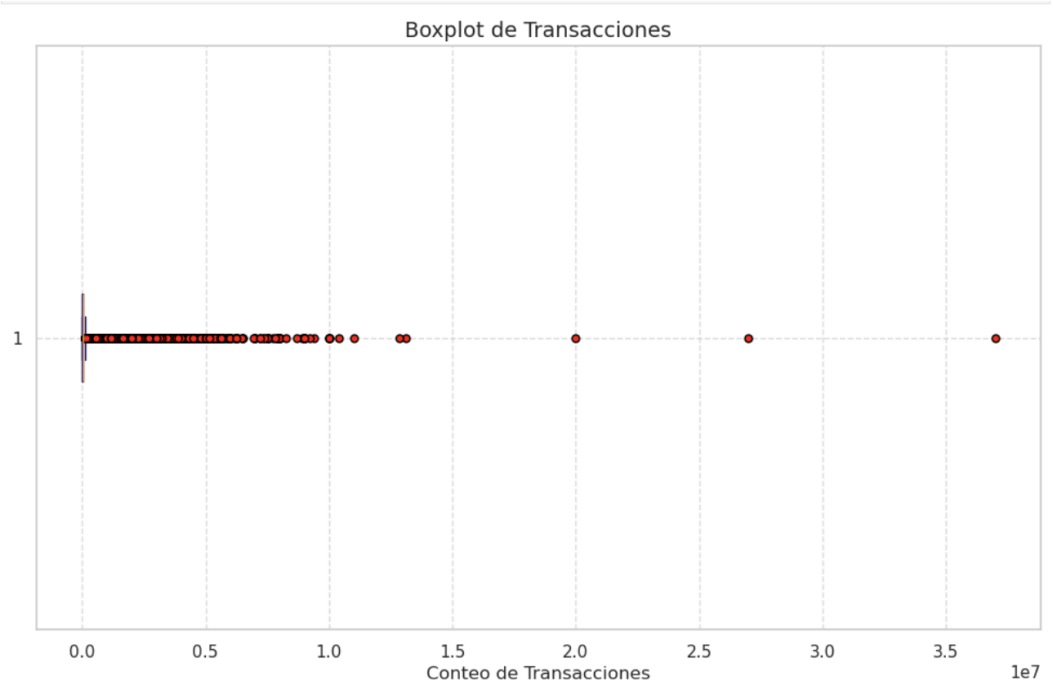
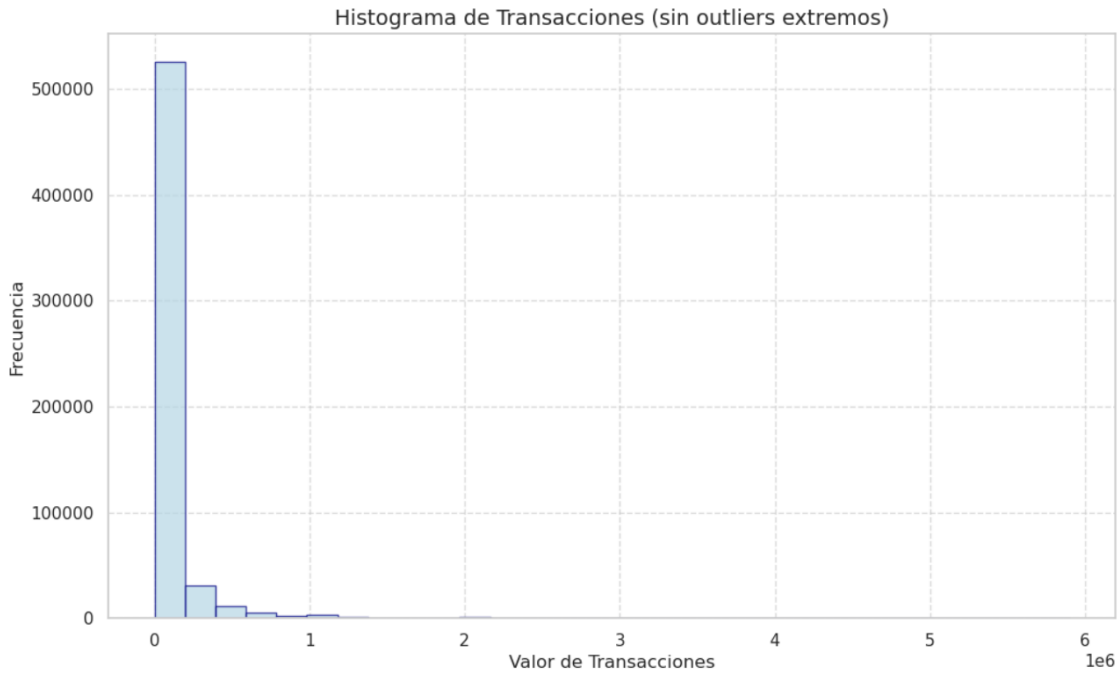


Figura 4

Distribución del valor de las transacciones - Histograma



El histograma muestra una distribución sesgada a la derecha, donde la mayoría de las contrapartidas tienen un bajo número de transacciones. La frecuencia disminuye rápidamente a medida que el número de transacciones aumenta. La mayoría de las contrapartidas tienen menos de 100 transacciones, con un pico claro en el rango más bajo (menos de 50 transacciones). A medida que el número de transacciones aumenta, la frecuencia de contrapartidas disminuye significativamente.

Para concluir sobre los dos gráficos anteriores, los cambios en el número de transacciones por contrapartida es alta, con algunas cantidad de envíos registrando un número excepcionalmente elevado de transacciones (datos atipicos), mientras que la mayoría tienen un número relativamente bajo. La concentración de transacciones se encuentra principalmente en el rango bajo, con pocas contrapartidas recibiendo muchas transacciones. Los datos atipicos identificados requieren un análisis detallado para comprender las razones detrás de estas anomalías, que podrían deberse a factores específicos del negocio o a irregularidades que deben abordarse. Estos gráficos proporcionan una visualización clara de la distribución de las transacciones, ofreciendo una base

sólida para tomar decisiones informadas y como punto de partida para realizar análisis más profundos.

Finalmente se desea identificar, mediante un mapa de calor y un histograma, el nivel transaccional por hora de manera agregada. La hora donde más se identificó transaccionalidad es las 19 horas con 43.481 registros y la de menor transaccionalidad con 1091 registros fue las 3 horas en dicho periodo de análisis, esto permite un acercamiento a la distribución de normalidad horaria de 9 horas hasta las 20 horas, siendo el viernes el día de la semana más transaccional como se evidencia en la siguiente tabla. Lo anterior se realiza con el ánimo de buscar patrones iniciales de perfiles transaccionales para complementar los ejercicios siguientes y lograr conclusiones prometedoras sobre el manejo del riesgo en entidades bancarias.

Figura 5

Transaccionalidad por hora - Mapa de calor

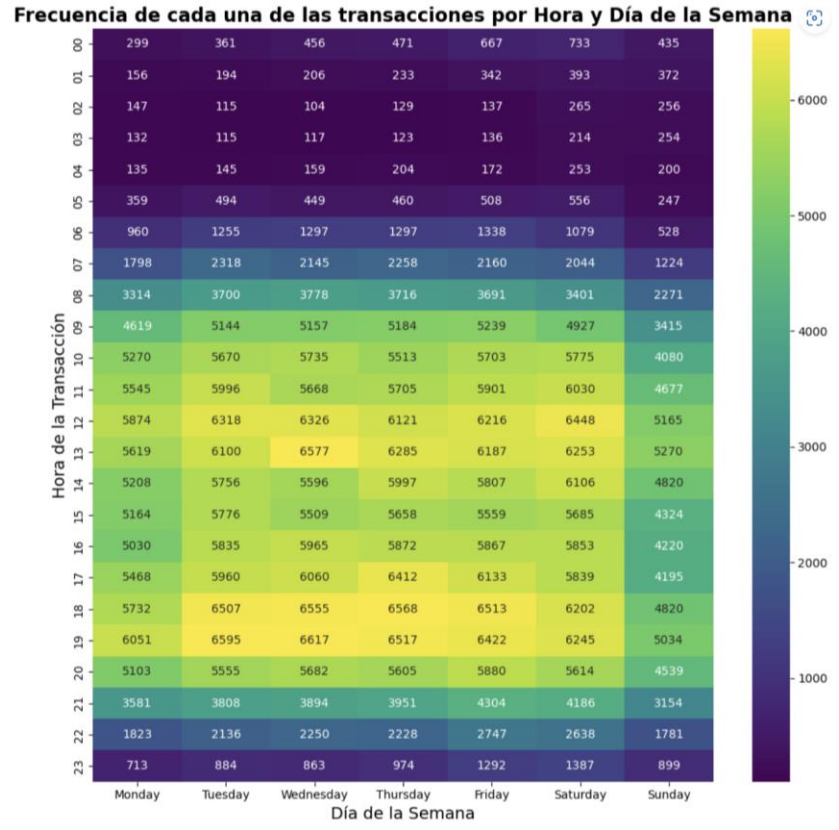


Figura 6

Distribución de transaccionalidad por hora - Histograma

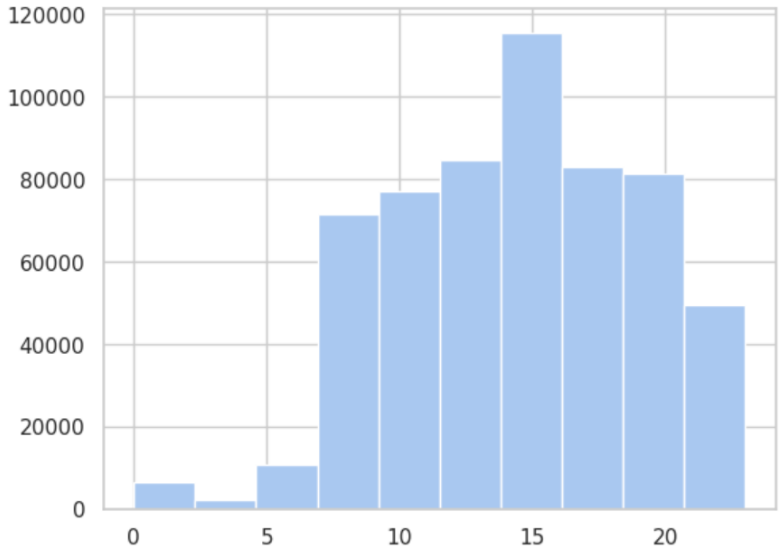


Tabla 2

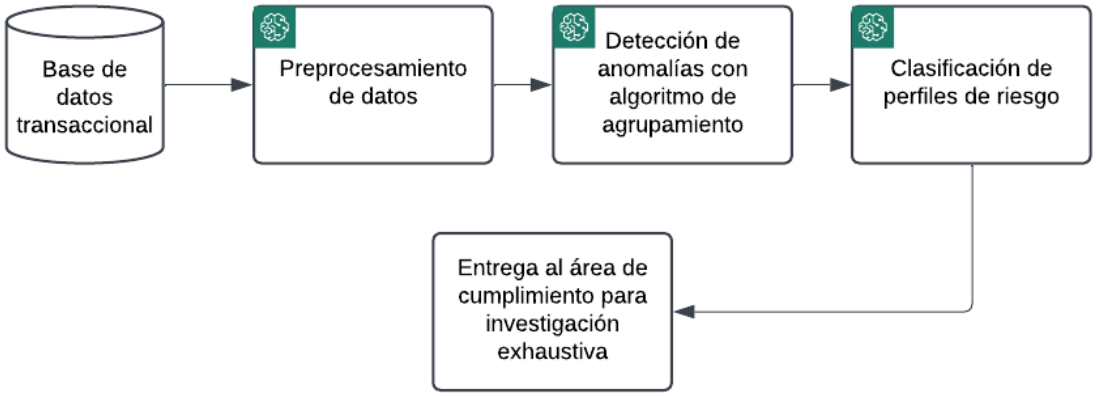
Transaccionalidad según el día de la semana.

Día de la Semana	Conteo de Transacciones	Porcentaje (%)
Lunes	78.100	1.40
Martes	86.737	14.89
Miércoles	87.165	14.96
Jueves	87.481	15.01
Viernes	88.921	**15.26**
Sábado	88.126	15.12
Domingo	66.180	11.36

4. Proceso de analítica

4.1. Pipeline Principal

Figura 7
Pipeline principal



4.2. Preprocesamiento

Durante el preprocesamiento, se aplicó StandardScaler definido como $z = \frac{\{x - \mu\}}{\sigma}$ donde x es el valor de cada característica, μ es la media de la característica y σ es la desviación estándar de la característica. Usada para estandarizar las variables, ajustando la media a cero y la desviación estándar a uno, lo que facilita el análisis y mejora el desempeño de los modelos. En la primera fase, al tratarse de un análisis no supervisado, no se consideró adicionar otros datos sintéticos. Sin embargo, en la segunda fase, el proyecto abordó un problema de clasificación donde una clase estaba significativamente desbalanceada. Para remediar este desbalance, se aplicó SMOTE (Synthetic Minority Over-sampling Technique) (Chawla et al., 2002) y se aplicó con la implementación desarrollada por imbalanced-learn (Lemaître et al., 2017), que generó nuevas muestras de la clase minoritaria, ayudando a mejorar el equilibrio y la representatividad de los datos en el modelo de clasificación.

4.3. Modelos

Para la identificación de comportamientos transaccionales anómalos en clientes, se implementaron una serie de modelos de agrupamiento. Primero, con K-Means, se ajustó el número óptimo de grupos utilizando el método del codo, el índice de la silueta, el índice de Calinski-Harabasz y el indicador de Davies-Bouldin. También se experimentó con agrupamiento espectral, evaluando el índice de la silueta para diferentes números de grupos y configurando affinity='nearest_neighbors' para captar relaciones locales entre puntos cercanos, lo que permite definir mejor la estructura de grupos en datos complejos. A continuación, se ajustó un modelo de *Fuzzy C-Means*, evaluando métricas como el Coeficiente de Partición (PC), el Coeficiente de Partición Modificado (MPC), el Coeficiente de Entropía Modificado (MPE), el índice Xie-Beni (XB), el índice Fukuyama-Sugeno (FS) y el índice Wu-Ling (WL), ajustando los centros de grupos hasta alcanzar una agrupación óptima. Finalmente, se ajustó un modelo de Isolation Forest, Liu *et al,* (2008) para la detección de átipicos, optimizando parámetros como *contamination* = 0.0101, *max_features* = 0.5, *max_samples* = 0.7 y *n_estimators* = 294 para separar eficientemente las observaciones anómalas.

Para la etapa de clasificación y mejora en la detección de usuarios con patrones anómalos, se implementaron modelos supervisados tales como los que se enuncian en la tabla 3, donde todos los modelos fueron optimizados mediante la librería **Optuna** (Akiba et al., 2019), que permitió iterar eficientemente sobre sus parámetros para maximizar las métricas objetivo.

Tabla 3
Iteraciones con algoritmos supervisados

Modelo	Remuestreo	Hiperparámetros
Random Forest	No	- max_depth = 29 - min_samples_split = 9 - n_estimators = 876
Random Forest (SMOTE)	Sí (SMOTE)	- max_depth = 47 - min_samples_leaf = 2 - min_samples_split = 11 - n_estimators = 487
HistGradientBoostingClassifier	No	- learning_rate = 0.0616 - max_depth = 18 - max_iter = 264
Support Vector Machine Classifier	No	- C = 0.0520 - kernel = 'linear'

4.4. Métricas

Para evaluar el desempeño de los modelos, se emplearon métricas específicas para cada tipo de análisis, adaptadas tanto al agrupamiento como a la clasificación.

En los modelos de agrupamiento, se aplicaron índices que miden la cohesión y separación entre los grupos:

- El **método del codo** evalúa la inercia (o suma de distancias cuadradas) dentro de los grupos; el punto en el que la reducción de inercia se estabiliza sugiere un número óptimo de grupos.
- El **índice de silueta** mide la coherencia de los grupos al comparar la distancia media de un punto con otros dentro del mismo grupo frente a su distancia a puntos en el grupo más cercano. Un valor alto indica buena separación.
- El **índice de Calinski-Harabasz** calcula la relación entre la dispersión dentro de los grupos y entre ellos, siendo un valor alto indicador de una estructura bien definida.
- El **índice de Davies-Bouldin** mide la dispersión dentro de los grupos en relación con la separación entre ellos; un valor bajo sugiere que los grupos están bien separados y son compactos.

En la fase de clasificación, se emplearon métricas de desempeño:

- La **precisión (precision)** mide la proporción de verdaderos positivos entre los positivos predichos, evaluando la exactitud de las predicciones positivas.
- La **sensibilidad (recall)** mide la proporción de verdaderos positivos entre todos los positivos reales, reflejando la capacidad del modelo para capturar todos los casos positivos.
- El **puntaje F1 (f1-score)** es la media armónica entre precisión y sensibilidad, ofreciendo un balance cuando existe un desbalance de clases.
- La **matriz de confusión** permite verificar el equilibrio entre clases, mostrando la cantidad de verdaderos y falsos positivos y negativos.

Además, se generó la **curva ROC-AUC**, que mide la capacidad de discriminación del modelo en la clasificación, donde un valor más cercano a 1 indica una excelente capacidad para distinguir entre clases. Todas estas métricas fueron calculadas usando funciones de *sklearn*, proporcionando una evaluación integral que combina el rigor técnico con el contexto de negocio.

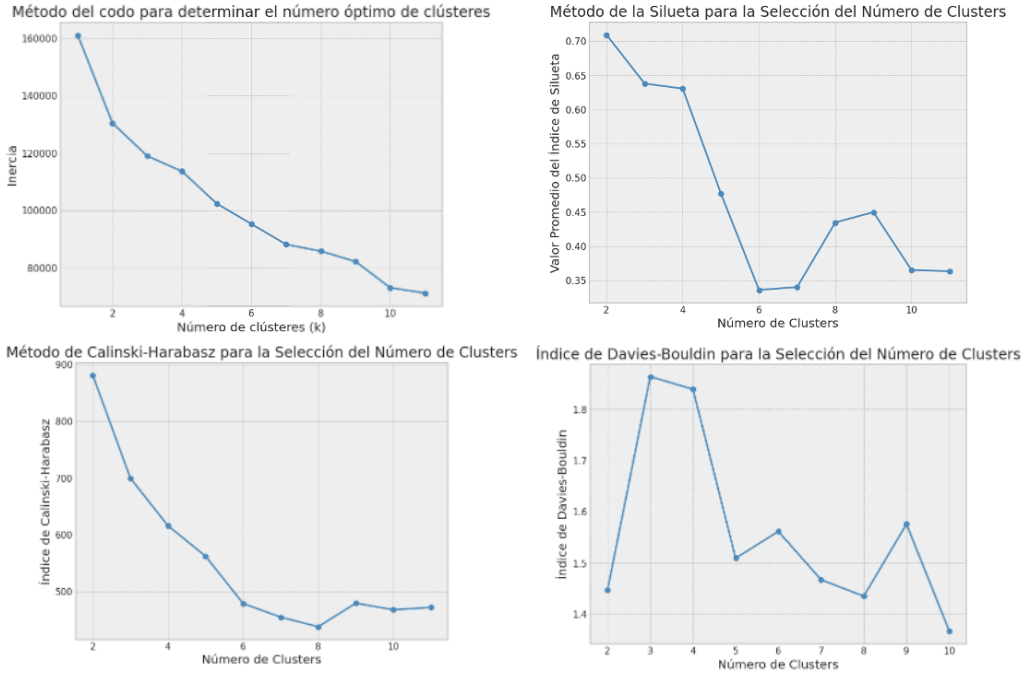
5. Metodología

5.1. Línea base

En la primera iteración, se ajustó un modelo K-means, precedido por un análisis exhaustivo de los índices de validación del número óptimo de grupos, incluyendo el método del codo, y los índices de silueta, Calinski-Harabasz y Davies-Bouldin. Cada uno de estos índices se utilizó para evaluar diferentes aspectos de la agrupación con el fin de seleccionar el número de grupos que mejor se ajustara a las características de los datos.

Figura 8

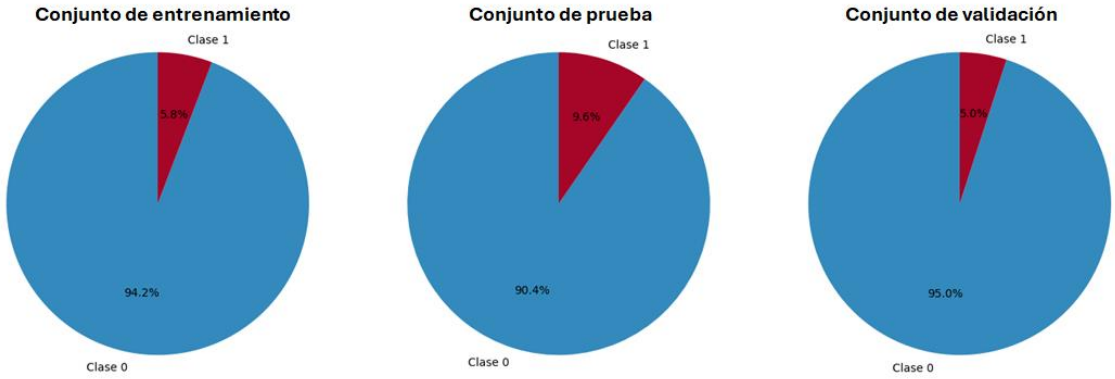
Índices para el número óptimo de grupos usando Kmeans - Gráficos de líneas



En esta primera iteración, se obtuvo un índice de silueta de 0.71 para $k=2$, un valor de Calinski-Harabasz de 881 (el más alto) para el mismo k, y un índice Davies-Bouldin de 1.44. Estos resultados indican que el número óptimo de grupos es **2**, lo cual se alinea con el objetivo principal de segmentar a los usuarios en dos categorías: aquellos con una transaccionalidad normal y aquellos con una transaccionalidad anómala.

Figura 9

Segmentación de usuarios con base en su transaccionalidad usando Kmeans - Gráficos de tortas



Aunque se identificó un grupo claramente minoritario, en futuras iteraciones se buscó mejorar la separación entre los grupos optimizando principalmente el índice de silueta y probando otros algoritmos especializados en la detección de anomalías. Esto tenía como objetivo lograr una segmentación más robusta de comportamientos anómalos, más allá de simplemente agrupar a los usuarios con transacciones por encima del promedio. Además, se consideró complementar esta agrupación mediante el ajuste de un modelo de clasificación, con el fin de fortalecer la identificación de usuarios con transaccionalidad anómala y aumentar la precisión en su detección.

5.2. Validación.

Para el entrenamiento de los algoritmos de agrupamiento, se realizó una partición inicial del conjunto de datos, destinando un 60% de los registros para el conjunto de entrenamiento y el 40% restante para pruebas y validación, dividiendo esta última proporción en partes iguales (20% para pruebas y 20% para validación externa o "fuera de muestra"). Este enfoque permitió evaluar el rendimiento de los modelos tanto en datos no vistos previamente como en datos completamente externos, ofreciendo una validación más robusta.

Además, se garantiza que la partición se realizará sobre los identificadores únicos (IDs) de cada usuario, evitando que la misma persona aparezca en diferentes subconjuntos. Este proceso fue fundamental para prevenir la fuga de información entre conjuntos, que podría sesgar los resultados del modelo y sobreestimar su desempeño. Esta estrategia de partición fue replicada en el

entrenamiento y la validación de los algoritmos de clasificación para mantener la coherencia en la metodología y asegurar una evaluación rigurosa de cada modelo en condiciones comparables.

5.3. Iteraciones y evolución

Tras la primera iteración con K-means, el objetivo fue mejorar la cohesión y separación de los grupos, incrementando el índice de silueta inicial. Se realizaron iteraciones con *Fuzzy C-Means*, experimentando con distintos valores de k y asignando cada observación al grupo de mayor pertenencia para una clasificación estricta. El mejor índice de silueta obtenido fue de 0.53 para k=2, lo que reflejó una agrupación menos homogénea que con K-means, mostrando que, aunque *Fuzzy C-Means* captura relaciones más complejas, en este caso no mejoró la estructura de grupos alcanzada inicialmente, dado las bajas métricas de concentración y compactabilidad tales como las de ***Calinski-Harabasz y Davies-Bouldin***. Este análisis ayudó a ajustar la estrategia de agrupación en iteraciones futuras y a buscar otras alternativas.

Tabla 4
Iteraciones con el algoritmo Fuzzy C-Means

Grupos	***Coeficiente de Silueta***	***Índice de Calinski-Harabasz***	***Índice de Davies-Bouldin***	***Conteo de Muestras***
2 Grupos	*0.532*	*699.479*	*1.793*	*grupo 0: 3055, grupo 1: 692*
3 Grupos	*0.266*	*481.768*	*2.407*	*grupo 0: 2378, grupo 1: 422, grupo 2: 947*
4 Grupos	*0.147*	*366.567*	*3.204*	*grupo 0: 1994, grupo 1: 329, grupo 2: 553, grupo 3: 871*

Seguidamente, aprovechando las ventajas del Isolation Forest para detección de anomalías, se optimizó su ajuste usando la librería Optuna (Akiba et al., 2019), iterando sobre parámetros clave como *n_estimators, max_samples, contamination* y *max_features* para maximizar la separación entre anomalías y datos regulares. Este proceso arrojó resultados satisfactorios, con un índice de silueta de 0.84 en entrenamiento, 0.83 en pruebas y un 0.92 en la validación externa, evidenciando así que el modelo logró una detección de anomalías consistente y robusta en distintos conjuntos de datos.

Tabla 5
Índice de la silueta con el algoritmo Isolation Forest

Set de datos	***Coeficiente de Silueta***
Entrenamiento	0.846
Test	0.834
Validación	0.926

Tras etiquetar las anomalías en los conjuntos de datos, se ajustó un modelo *RandomForestClassifier* sin aumentar de forma sintética las muestras de la clase desbalanceada. Mediante 100 iteraciones con Optuna (Akiba et al., 2019), se optimizaron hiperparámetros clave como 'n_estimators', 'max_depth' y 'class_weight', maximizando el puntaje F1 por su enfoque en balancear precisión y sensibilidad.

Aunque se obtuvo un puntaje F1 de 0.81 y un AUC de 0.83 en prueba, la matriz de confusión reveló importantes limitaciones: varios casos de anomalías fueron clasificados erróneamente como usuarios normales. En validación, el desempeño decayó aún más (f1-score = 0.7), destacando problemas de consistencia en la detección de la clase minoritaria.

El Brier Score fue bajo (0.0065), reflejando buena calibración de probabilidades, pero no logró compensar la alta tasa de falsos negativos. Esto evidenció la necesidad de mejorar, explorando estrategias como técnicas de balanceo o ajustes en los pesos de clase para iteraciones posteriores.

En una iteración posterior, se utilizó **SMOTE** para balancear las clases mediante el aumento sintético de las muestras minoritarias. A partir de este nuevo conjunto de datos, se optimizó nuevamente el **RandomForestClassifier** con Optuna (Akiba et al., 2019). Sin embargo, los resultados empeoraron considerablemente: en el conjunto de prueba se obtuvo un **AUC** de 0.71 y un **puntaje F1** de 0.6, mientras que en el conjunto de validación el desempeño fue aún más bajo, con un **AUC** de 0.6 y un **puntaje F1** de 0.333.

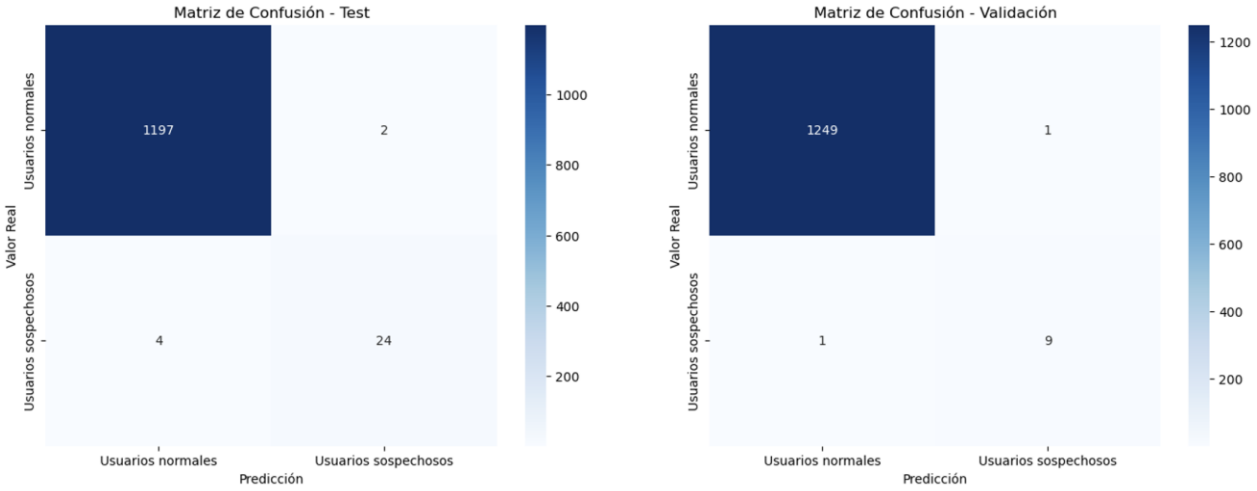
Estos resultados evidencian que, aunque el balance con SMOTE aumentó la representatividad de la clase minoritaria, el modelo no logró generalizar correctamente. La baja sensibilidad en validación, reflejada por el pobre puntaje F1, subraya la necesidad de explorar otros algoritmos más robustos para este tipo de problemas, como un **SVM Classifier**, que podría manejar mejor las complejidades de las fronteras entre clases.

Finalmente, se optimizó un algoritmo **Support Vector Classifier (SVC)** utilizando **100 iteraciones con Optuna (Akiba et al., 2019)**, explorando los valores para **C** y los diferentes tipos de **kernel** (*linear*, *poly*, *rbf* y *sigmoid*). Este modelo logró un notable desempeño:

- En el **conjunto de prueba**, se alcanzó un **AUC** de 0.93 y un **puntaje F1** de 0.88.
- En el **conjunto de validación**, los resultados fueron incluso mejores, con un **AUC** de 0.94 y un **puntaje F1** de 0.9.

Las matrices de confusión reflejan esta mejora, mostrando una reducción significativa en los falsos negativos y un equilibrio más adecuado entre precisión y sensibilidad. Este comportamiento evidencia que el **SVC**, con una configuración óptima de hiperparámetros, es capaz de capturar de manera eficiente las anomalías en los datos, superando a las iteraciones previas con Random Forest.

Figura 10
Matrices de Confusión SVM Classifier

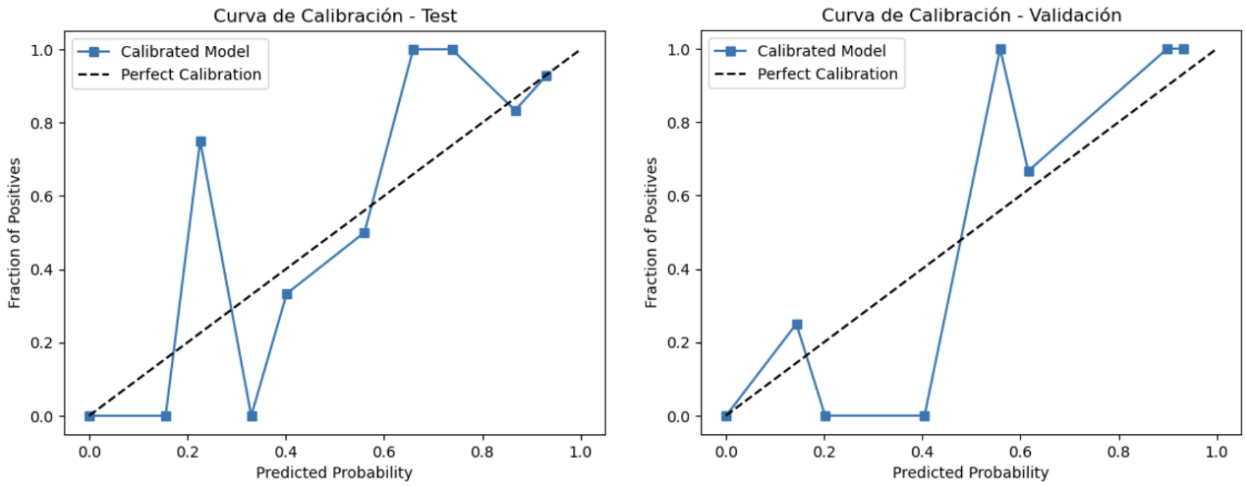


Con el fin de mejorar la interpretabilidad de las probabilidades predichas, se procedió a calibrar el modelo utilizando el algoritmo de regresión isotonica.

Las **curvas de calibración** obtenidas para los conjuntos de prueba y validación muestran que:

- En el conjunto de **prueba**, el modelo presenta una buena alineación en las probabilidades medias superiores al 0.5, lo que indica una correcta calibración en esas regiones. Sin embargo, en los valores bajos de probabilidad, existe una mayor discrepancia con respecto a la línea ideal de calibración (dashed line).
- En el conjunto de **validación**, el comportamiento es similar, aunque las predicciones parecen estar mejor calibradas en general.

Figura 11
Curvas de calibración - SVM Classifier



Estos resultados sugieren que el modelo calibrado es más confiable en sus predicciones para rangos altos de probabilidad, mientras que las probabilidades bajas podrían requerir ajustes adicionales o mayor cantidad de datos para perfeccionar la calibración. Esto refuerza la utilidad del ajuste isotónico en este escenario, particularmente para aplicaciones críticas donde las probabilidades deben reflejar el riesgo real.

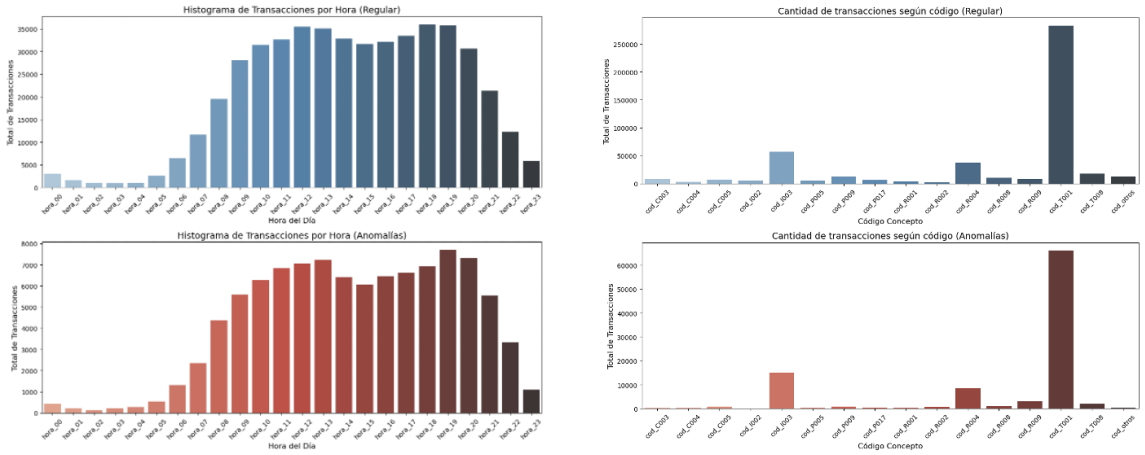
6. Resultados y discusión

El proceso de agrupamiento permitió segmentar efectivamente las anomalías utilizando el algoritmo Isolation Forest, obteniendo un desempeño destacado con índices de silueta de 0.85 en el conjunto de entrenamiento, 0.83 en el de prueba y 0.92 en validación (Figura 9). Esto indica una separación clara entre los grupos identificados.

Al analizar las transacciones, se observó que las distribuciones de frecuencia según las horas del día y los códigos de concepto no presentaron diferencias significativas entre los usuarios etiquetados como "regulares" y aquellos clasificados como "anómalos". Tal como se ilustra en los histogramas (Figura 12), los patrones horarios son consistentes, con picos de actividad entre las 10:00 y las 16:00 horas en ambos casos. Asimismo, los códigos más comunes corresponden a transacciones entre usuarios de la compañía.

Figura 12

Histogramas de transacciones según horas y códigos de conceptos

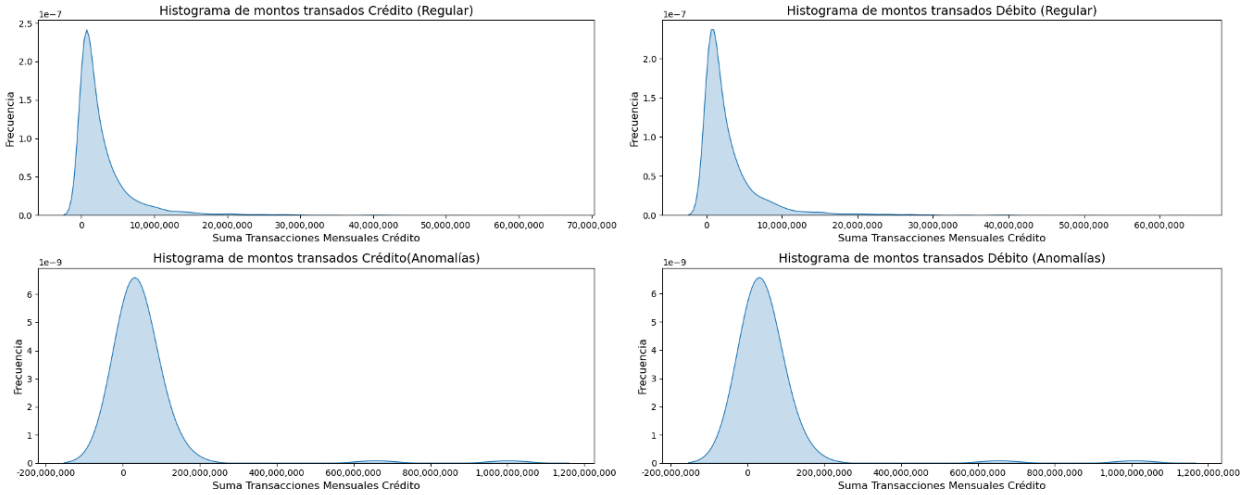


Sin embargo, al analizar los montos de las transacciones (débito y crédito), sí se identificaron diferencias significativas entre ambos grupos, los clasificados como regulares muestran una distribución más estable y uniforme, lo que sugiere un comportamiento predecible y controlado en las transacciones habituales. En cambio, las anomalías reflejan una mayor dispersión y extremos inusuales, lo que podría indicar errores, eventos excepcionales o comportamientos fuera de lo común.

Cuando comparamos las transacciones de crédito con las de débito, la diferencia más notable está en la amplitud de los montos extremos: las transacciones de crédito tienden a tener una mayor variabilidad, lo que podría deberse a la naturaleza de este tipo de operaciones, que suelen manejar valores más altos o menos frecuentes en comparación con el débito.

Figura 13

Distribuciones de montos transados (Débito y Crédito) según segmentación de usuarios



Como paso siguiente en el análisis, se desarrolló un modelo de clasificación para predecir la probabilidad de que un usuario sea clasificado como sospechoso, entrenado utilizando los resultados de la detección de anomalías no supervisadas. A partir de este modelo, se definieron los siguientes rangos para el perfil de riesgo:

- **Riesgo Bajo:** 0 a 0.25
- **Riesgo Moderado:** 0.26 a 0.50
- **Riesgo Alto:** 0.51 a 0.75
- **Riesgo Crítico:** 0.76 a 1

Posteriormente, se aplicó la predicción a los conjuntos de prueba y validación, asegurando que los usuarios de estos conjuntos no hubieran sido incluidos en el entrenamiento del modelo. Los resultados de la clasificación por perfil de riesgo fueron los siguientes:

- **Riesgo Bajo:** 2,442 usuarios
- **Riesgo Moderado:** 8 usuarios
- **Riesgo Alto:** 12 usuarios

- **Riesgo Crítico:** 25 usuarios

Estos resultados ofrecen una distribución clara de los perfiles de riesgo, permitiendo identificar un grupo reducido de usuarios con mayor probabilidad de presentar comportamientos sospechosos.

Esta información generada se pondría a disposición del área de cumplimiento, la cual es la responsable de llevar a cabo las investigaciones necesarias sobre los usuarios clasificados en los niveles de riesgo alto y crítico. Dichas investigaciones buscarán determinar si estos perfiles presentan comportamientos asociados con riesgos de Lavado de Activos y Financiación del Terrorismo (LA/FT). Esto permitirá enfocar los esfuerzos en los casos más relevantes y optimizar los recursos destinados a la prevención y mitigación de riesgos.

6.1. Métricas

Tabla 6

Métricas para el modelo de agrupamiento entrenado

Proceso de Aegrupamiento		
Mejor Modelo	**Métricas evaluadas**	**Resultados**
IsolationForest(contamination=0.01 1046, max_samples = 0.7, n_stimators = 110, random_state = 42)	Índice de la Silueta	• Train: 0.85 • Test: 0.83 • Validation: 0.93

Tabla 7

Métricas para el modelo de Clasificación entrenado

Proceso de Clasificación		
Mejor Modelo	**Métricas evaluadas**	**Resultados**
SVC (C = 0.047665, kernel = ‘linear’, probability = True, random_state = 42)	puntaje F1	0.88
	AUC	0.92
	Brier Score	0.004

6.2. Evaluación cualitativa

Desde el punto de vista técnico, los resultados obtenidos son satisfactorios. Se logró desarrollar un modelo de *agrupamiento* con un buen desempeño para separar anomalías, complementado con un modelo de clasificación preciso y calibrado para predecir estos comportamientos anómalos. Sin embargo, se identifican oportunidades de mejora, particularmente en la calidad y cantidad de información disponible sobre los clientes.

Observaciones y puntos de mejora:

- **Limitaciones de las variables utilizadas:** La información actual utilizada para entrenar los modelos podría ser enriquecida con más variables relevantes y una mayor profundidad histórica. Esto ayudaría a capturar patrones más complejos y representativos de los comportamientos sospechosos. Actualmente, existe el riesgo de clasificar como sospechosos a clientes que simplemente presentan una transaccionalidad elevada, sin que ello necesariamente implique un comportamiento fraudulento**.**
- **Relación entre métricas de ML y métricas de negocio:** Aunque el índice de la silueta obtenido en el modelo de *agrupamiento* refleja una adecuada separación entre grupos, su relación con las métricas de negocio, como la reducción de riesgos asociados al Lavado de Activos y Financiación del Terrorismo (LA/FT), aún puede ser fortalecida mediante una mayor personalización en la segmentación.
- **Riesgo de sobreajuste y subajuste:** No se observan evidencias de *overfitting* en los resultados, ya que el modelo logró generalizar correctamente al predecir sobre datos de prueba y validación. Sin embargo, un posible *subajuste* puede surgir si no se incorporan variables clave que expliquen mejor los comportamientos anómalos, dejando espacio para mejorar el poder predictivo.

Utilidad de los resultados:

En términos de negocio, la metodología propuesta aporta valor significativo:

- Facilita el enfoque en segmentos más específicos y relevantes, optimizando los esfuerzos del área de cumplimiento.
- Mejora la eficiencia en la identificación y mitigación de riesgos asociados a LA/FT al priorizar casos con mayores probabilidades de ser fraudulentos.

En conclusión, los resultados obtenidos son útiles y representan un avance importante en la mitigación de riesgos. No obstante, incorporar más información sobre los clientes y ajustar los modelos con datos más robustos podría incrementar significativamente la precisión y la utilidad práctica del sistema, permitiendo tomar decisiones más informadas y efectivas en la gestión del riesgo**.**

6.3. Consideraciones de producción.

Para desplegar los modelos en AWS, se recomienda:

6.3.1. Supervisión del modelo de *agrupamiento*:

- Monitorear la proporción de anomalías detectadas usando *Amazon CloudWatch Logs*.
- Detectar *data drift* con *SageMaker Model Monitor*.
- Realizar análisis periódicos con tableros de control para garantizar consistencia en los datos de entrada.

6.3.2. Supervisión del modelo de clasificación:

- Evaluar continuamente precisión, recall y *Brier score* mediante *SageMaker Clarify*.
- Detectar degradación o *concept drift* analizando predicciones recientes frente a las históricas.
- Reentrenar si se observa deterioro en el desempeño.

6.3.3. Despliegue técnico:

- Publicar ambos modelos como endpoints en *SageMaker* y orquestar con Apache Airflow para evitar manualidades tanto en la ejecución de las ETL como en la ejecución de los modelos.

- Configurar un flujo CI/CD con *AWS CodePipeline* para actualizaciones automáticas.

6.3.4. Alertas y visualización:

- Crear tableros de control para riesgos y tendencias, y configurar notificaciones críticas con *SNS* o *EventBridge*.

7. Conclusiones

La detección de comportamientos atípicos en el campo financiero sigue siendo un reto importante de monitorear y se deben asumir múltiples retos dadas las tendencias en el sector financiero digital. Por ello, la solución desarrollada ofrece un marco de trabajo eficaz en la detección de usuarios con comportamientos sospechosos, optimizando esfuerzos del área de cumplimiento, pues la segmentación obtenida permite concentrar los recursos en perfiles de alto riesgo, mejorando la eficiencia operativa del área de este modo reducir el tiempo invertido en investigaciones de actividades ilícitas.

El sistema propuesto no solo contribuye a la mitigación de riesgos asociados al lavado de activos y la financiación del terrorismo, sino que también refuerza el cumplimiento normativo, protegiendo la reputación de Nequi Compañia de Financiamiento S.a y fortaleciendo la confianza en su operación. Además, la metodología adoptada tiene potencial de escalabilidad y adaptabilidad para integrar nuevos datos asegurando su uso en un largo plazo.

Para trabajos futuros, se sugiere incorporar nuevas variables que reflejen cambios históricos en la actividad de los clientes, lo que podría mejorar la precisión en la detección de anomalías. Además, se recomienda explorar la identificación de comunidades anómalas mediante análisis de grafos, así como el uso de autoencoders para modelar patrones complejos y no lineales en los datos, ampliando las capacidades del sistema para adaptarse a comportamientos emergentes.

Referencias:

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, julio). Optuna: A next-generation hyperparameter optimization framework. En *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623-2631).
- Buchanan, B. (2004). Money laundering—a global obstacle. *Research in International Business and Finance*, 18(1), 115-127.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
- Chen, Y. T., & Mathe, J. (2011). Fuzzy computing applications for anti-money laundering and distributed storage system load monitoring. En *World Conference on Soft Computing*.
- Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: A review. *Knowledge and Information Systems*, 57, 245-285.
- Dumitrescu, B., Băltoiu, A., & Budulan, Ş. (2022). Anomaly detection in graphs of bank transactions for anti-money laundering applications. *IEEE Access*, 10, 47699-47714.
- Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of machine learning-based k-means clustering for financial fraud detection. *Academic Journal of Science and Technology*, 10(1), 33-39.
- Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17), 1-5.
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, 413–422. IEEE Xplore.

- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. En *2008 Eighth IEEE International Conference on Data Mining* (pp. 413-422). IEEE.
- Masciandaro, D. (2013). Is the anti-money laundering compliance convenient?: International capital flows and stigma effect in Latin America: The case of Paraguay.
- Powell, J. H. (2013). Anti-Money Laundering and the Banking Secrecy Act, *Board of Governors of the Reserve System, Committee on Banking, Housing and Urban Affairs, U.S. Senate*, Washington D.C., marzo 7. mimeo.
- Unidad de Información y Análisis Financiero (UIAF). (2014). *La dimensión económica del lavado de activos*. UIAF. https://www.uiaf.gov.co/?idcategoria=20493&download=Y

Anexos

Anexo 1. Variables en dataset utilizado en la modelación.

Columna	Valores Únicos	Definición
dia_transaccion	225	Dia transacción en dia-mes-año
tipo_identificacion	3	Tipo de identificación
codigo_producto	2	Sí es una cuenta de bajo monto o no
numero_producto	993	Indicadora
hora_transaccion	70877	Hora en la que se efectúa la transacción
valor_transaccion	24837	Monto transado
naturaleza_transaccion	2	identificador si es de naturaleza crédito o debito
codigo_concepto	71	Código asociado a la descripción del movimiento
concepto_desc	28	Descripción del movimiento
canal_desc	8	Canal de envío/recepción de dinero
entidad_contraparte	29	Entidad que recibe la transacción
numero_cuenta_contraparte	130200	identificador que recibe el monto enviado
tipo_identificacion_contraparte	4	Tipo de documento del que recibe
numero_celular	1047	Número de celular donde se realiza el movimiento
clase	6	Códigos para ajustes contables por tipo de entrada o salida
codigo_banco_destino	17	Código del banco al que se envia el dinero
tipo_cuenta_destinon	4	
numero_cuenta_destino	192	Número de cuenta donde se envia el monto
codigo_servicio	234	
year	2	Partición por año ingestión
month	8	Partición por mes ingestión
day	31	Partición por día ingestión
rn3	701	Variable indicadora de listas de control o no

desc_concepto	52	Concepto por el que se envía o se marca el movimiento financiero