

Proyecto AML simulado

Paola Espinoza Hernández

Resumen

En este proyecto se construye un conjunto de datos sintético para analizar reglas de monitoreo transaccional orientadas a la prevención del lavado de dinero (AML). Se simulan 1000 clientes distribuidos entre las siete provincias de Costa Rica y los segmentos Personal, PYME y Corporativo, incorporando la condición de Persona Expuesta Políticamente (PEP) y niveles de ingreso mensual diferenciados por segmento y por condición PEP. A partir de estos clientes, se generan transacciones para el año 2024, con montos modelados mediante distribuciones lognormales vinculadas al salario. A partir de estos datos, se define un conjunto de reglas que marcan transacciones como sospechosas cuando se exceden ciertos umbrales de monto diario, frecuencia diaria o montos por segmento, incluyendo umbrales específicos para cuentas personales PEP. Las reglas se aplican con un orden de prioridad y el resultado se resume en una variable de regla principal, lo que permite calcular tasas de alerta globales y por segmento, así como la distribución de alertas por tipo de regla. El análisis muestra que se generan 112 alertas en el periodo simulado, concentradas en el segmento personal y dominadas por la regla de monto máximo por segmento, seguida por la regla basada en el salario mensual, mientras que la regla asociada a PEP aporta un número menor de casos. Estos resultados ilustran cómo el diseño de umbrales y reglas por segmento influye directamente en el número y composición de las alertas producidas.

Palabras clave: AML, lavado de dinero

Introducción

El presente proyecto utiliza datos simulados, disponibles en el [repositorio del proyecto](#), de modo que dichos datos son de uso exclusivamente académico. El proyecto se enfoca en la creación de diversas reglas para la detección de patrones sospechosos, en el contexto de la prevención del lavado de dinero a través del monitoreo de transacciones.

El enfoque consiste en construir un conjunto de datos sintético que permita controlar los supuestos de segmentación y comportamiento transaccional, y estudiar cómo distintas configuraciones de reglas afectan el volumen y la com-

posición de las alertas generadas. Para ello se simulan 1000 clientes asignados aleatoriamente a las siete provincias de Costa Rica y a los segmentos Personal, PYME y Corporativo, incorporando un subconjunto de Personas Expuestas Políticamente (PEP) y niveles de ingreso mensual diferenciados por segmento y por condición PEP. Posteriormente se generan transacciones para el año 2024, definiendo para cada una un canal, un tipo de operación, la moneda, una fecha y un monto aleatorio relacionado con el salario del cliente.

Una vez creado el conjunto de datos, se define un modelo de reglas AML basado en umbrales de monto diario, frecuencia diaria de transacciones y montos máximos por segmento y condición PEP. Estas reglas se aplican con un orden de prioridad para obtener, por transacción, la activación de alertas y la regla principal asociada. Finalmente, se calculan indicadores agregados por segmento y por tipo de regla, y se analizan los resultados en términos del número total de alertas, su distribución entre segmentos y la contribución relativa de cada regla, lo que permite evaluar el efecto de los parámetros definidos en la simulación sobre el patrón de alertas observado.

Metodología

Simulación de datos

Clientes

Se simularon 1000 clientes, asignados aleatoriamente mediante la función `sample` entre las siete provincias de Costa Rica y los segmentos Personal, PYME y Corporativo. Adicionalmente, se incorpora la condición de Persona Expuesta Políticamente (PEP) para un subconjunto de clientes, dada su relevancia en el contexto AML. La asignación de salarios se realiza mediante valores aleatorios provenientes de distribuciones lognormales con medias de 600 000, 1 200 000 y 4 000 000, de acuerdo con el segmento al que pertenece cada cliente. En el caso de los clientes PEP, el ingreso mensual se obtiene multiplicando el salario base por un factor de 1.8.

Transacciones

Para la simulación de las transacciones se asigna, a cada operación, un canal entre “Cajero”, “Ventanilla”, “Internet” y “SINPE”; un tipo de transacción entre “Depósito”, “Retiro”, “Transferencia” y “Pago”; la moneda de la operación (dólares o colones); y una fecha dentro del año 2024. Los montos se generan a partir de distribuciones lognormales cuya media corresponde aproximadamente a una décima parte del salario mensual del cliente asociado.

Modelo AML

Una vez creados los datos, se definen los siguientes umbrales para marcar una transacción como sospechosa:

- Montos diarios mayores o iguales al salario mensual.
- Montos diarios mayores o iguales a 3 000 000.
- Cinco o más transacciones en un mismo día.
- Montos mayores o iguales al umbral por segmento: 800 000 para cuentas personales, 1 000 000 para PEP, 1 500 000 para cuentas PYME y 5 000 000 para cuentas corporativas.

Adicionalmente, se genera un archivo `.csv` que incluye la variable `id_regla_principal`, en la cual se establece un orden de prioridades entre las reglas: primero el umbral por monto, luego el umbral por salario y, finalmente, el umbral por frecuencia (monto y frecuencia diarios). En consecuencia, si una transacción sobrepasa el umbral de monto, se clasifica únicamente por esa regla, aunque también exceda el umbral de salario. Sin embargo, el archivo `alertas.csv` permite identificar cuántos y cuáles de los umbrales se sobrepasan en cada transacción marcada.

Resultados

En total se identificaron 112 transacciones que activaron al menos una de las reglas de alerta definidas. La tabla de indicadores globales muestra el número total de operaciones analizadas, el número de alertas detectadas y la tasa de alertas resultante sobre el total de transacciones del periodo.

Table 1: Distribución de alertas por segmento

segmento	transacciones	alertas	tasa de alertas
Personal	14147	88	0.0062204
Corporativo	1850	10	0.0054054
PYME	4003	14	0.0034974
Total	20000	112	0.0056000

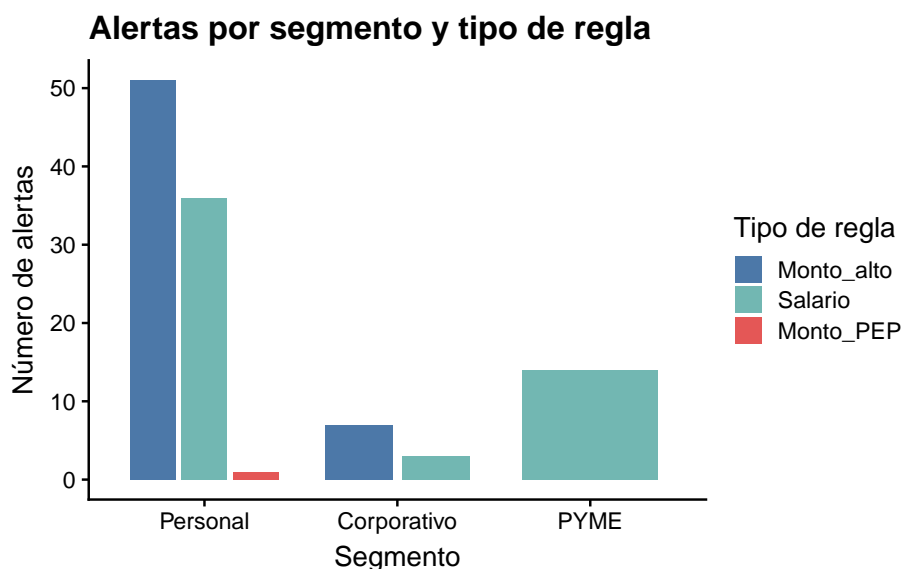
Al desagregar por segmento, la tabla de resultados indica que las cuentas personales registran el mayor número de alertas y la mayor tasa de alertas en relación con la cantidad de transacciones que realizan. Los segmentos corporativo y PYME presentan tanto un menor volumen de transacciones con alerta como tasas de alerta inferiores a las observadas en el segmento personal. La fila “Total” permite comparar estas tasas segmentadas con el comportamiento agregado de la cartera.

Table 2: Distribución de alertas por regla principal

regla principal	alertas	proporción
Monto_alto	58	0.5178571
Salario	53	0.4732143

Monto_PEP	1	0.0089286
-----------	---	-----------

El resumen por regla principal muestra que la mayor parte de las alertas se explica por la regla asociada al umbral de monto máximo por segmento, seguida por la regla basada en el salario del cliente. La regla Monto_PEP aporta un número reducido de alertas en comparación con las anteriores, consistente con el hecho de que solo puede activarse para clientes personales marcados como PEP.



El gráfico de alertas por segmento y tipo de regla confirma estos patrones: en el segmento personal se observan alertas para las tres reglas consideradas, mientras que en el segmento corporativo solo se activan las reglas de Monto_alto y Salario, y en el segmento PYME las alertas corresponden únicamente a la regla basada en el salario. En conjunto, estos resultados indican que, bajo los supuestos de simulación utilizados, la combinación de segmento y regla define perfiles de alerta diferenciados entre cuentas personales, corporativas y PYME.

Conclusiones y Recomendaciones

Los resultados muestran que el sistema de reglas genera 112 alertas en el periodo analizado. La comparación entre el total y los segmentos, indica que las cuentas personales concentran tanto el mayor volumen de alertas como la mayor tasa de alerta respecto del número de transacciones que realizan. Además, al descomponer las alertas por tipo de regla, se observa que la mayor parte se origina en la regla asociada al umbral de monto máximo por segmento, seguida por la regla basada en el ingreso mensual del cliente.

La distribución conjunta por segmento y regla confirma que las cuentas personales son el único segmento en el que se activan las tres reglas consideradas, mientras que en el segmento corporativo solo se observan alertas por Monto_alto y Salario, y en el segmento PYME únicamente por la regla de Salario. Esto indica que la combinación de segmento y tipo de regla determina patrones de alerta diferenciados dentro de la cartera. Por tanto, estos resultados indican que el diseño de los umbrales y de las reglas por segmento tiene un efecto directo sobre el volumen y la composición de las alertas generadas.

A partir de estos resultados, un siguiente paso natural sería revisar la calibración de las reglas y umbrales utilizados en la simulación, en particular la regla de monto máximo por segmento y la regla basada en el salario, dado que explican la mayor parte de las alertas detectadas. Además, podría estudiarse con mayor detalle el comportamiento de la regla Monto_PEP, verificando bajo qué configuraciones de parámetros comienza a generar un nivel de alertas más relevante en relación con su población objetivo. Sobre esta base, el ejercicio puede extenderse mediante nuevas rondas de simulación con distintos supuestos, la incorporación de variables adicionales como historial transaccional o características geográficas.