

MÉTODOS DE CLASIFICACIÓN PARA LA SEGMENTACIÓN – SARLAFT

Yuly Andrea Lizarralde

Índice

1. INTRODUCCIÓN

HSDHSGDHAGD

El objetivo de este trabajo es hacer uso de la teoría de métodos de clasificación para realizar un modelo aplicado a la gestión de riesgo que permita efectuar la segmentación de clientes vinculados a una entidad financiera. Esto tomando en cuenta que la Superintendencia Financiera de Colombia considera que los clientes o usuarios son unos de los agentes generadores del riesgo de lavado de activos y financiación del terrorismo (en adelante LA/FT). La segmentación, según se define en la circular externa 022 del 2007, es el proceso por medio del cual se lleva a cabo la separación de elementos en grupos homogéneos al interior de ellos y heterogéneos entre ellos. La separación se fundamenta en el reconocimiento de diferencias significativas en sus características (variables de segmentación). Las variables de segmentación que se tendrán en cuenta para los clientes son: actividad económica, volumen o frecuencia de sus transacciones, monto de ingresos y egresos y patrimonio, etc.

Para iniciar veremos las definiciones de lavado de activos y financiación del terrorismo que se encuentran estipuladas en la ley colombiana, que es SAR-LAFT y que deben hacer las entidades vigiladas para implementarlo. Después se darán las definiciones pertinentes sobre los principales métodos de clasificación y se intentará mostrar de que manera se puede aplicar la teoría de un método de clasificación seleccionado a la segmentación de clientes que deben efectuar las entidades vigiladas por la superintendencia financiera para identificar el riesgo de LA/FT

2. Preliminares

2.1. Lavado de Activos

Lavar activos es tratar de dar apariencia de legalidad a recursos de origen ilícito y constituye un serio problema en nuestro país, principalmente por que se encuentra ligado a actividades delictivas como el narcotráfico, tráfico de migrantes, trata de personas, extorsión, enriquecimiento ilícito, secuestro, tráfico de armas, corrupción y delitos contra el sistema financiero.

El artículo 323 del código penal define el delito de lavado de activos así:

el que adquiera, resguarde, invierta, transporte, transforme, custodie o administre bienes que tengan su origen mediano o inmediato en actividades de tráfico de migrantes, trata de personas, extorsión, enriquecimiento ilícito, secuestro extorsivo, rebelión, tráfico de armas, financiación del terrorismo y administración de recursos relacionados con actividades terroristas, tráfico de drogas tóxicas, estupefacientes o sustancias sicotrópicas, delitos contra el sistema financiero, delitos contra la administración pública, o vinculados con el producto de delitos ejecutados bajo concierto para delinquir, o les dé a los bienes provenientes de dichas actividades apariencia de legalidad o los legalice, oculte o encubra la verdadera naturaleza, origen, ubicación, destino, movimiento o derecho sobre tales bienes o realice cualquier otro acto para ocultar o encubrir su origen ilícito. (art.323, ley 599 del 2000)

2.2. Financiación del Terrorismo

En Colombia es un delito financiar el terrorismo o administrar recursos relacionados con actividades terroristas.

Se entiende por financiación del terrorismo: *El que directa o indirectamente provea, recolecte, entregue, reciba, administre, aporte, custodie o guarde fondos, bienes o recursos, o realice cualquier otro acto que promueva, organice, apoye, mantenga, financie o sostenga económicamente a grupos armados al margen de la ley o a sus integrantes, o a grupos terroristas nacionales o extranjeros, o a terroristas nacionales o extranjeros, o a actividades terroristas. (art. 345, ley 599 del 2000).*

A diferencia del lavado de activos, en la financiación del terrorismo el origen de los recursos puede ser lícito

2.3. SARLAFT

Es el sistema de administración que deben implementar las entidades vigiladas por la Superintendencia Financiera de Colombia para protegerse frente al riesgo de LA/FT y se instrumenta a través de las etapas y elementos que lo integran. (Norma 022 de SuperFinanciera de Colombia)

Según la Superintendencia Financiera de Colombia para identificar el riesgo de LA/FT las entidades vigiladas deben como mínimo:

1. Establecer las metodologías para la segmentación de los factores de riesgo.
2. Con base en las metodologías establecidas, segmentar los factores de riesgo.
3. Establecer las metodologías para la identificación del riesgo de LA/FT y sus riesgos asociados respecto de cada uno de los factores de riesgo.

4. Con base en las metodologías establecidas en desarrollo del literal anterior, identificar las formas a través de las cuales se puede presentar el riesgo de LA/FT.

Clientes o Usuarios representan uno de los factores de riesgo que para los efectos del SARLAFT deben tener en cuenta las entidades vigiladas. Según la Superintendencia Financiera de Colombia, el cliente es toda persona natural o jurídica con la cuál la entidad establece y mantiene una relación contractual o legal para el suministro de cualquier producto propio de su actividad.

Tomando esto en consideración, es el propósito de este trabajo exponer una metodología para la segmentación de clientes haciendo uso de la teoría de métodos de clasificación.

3. MÉTODOS DE CLASIFICACIÓN

Escribir resumen sobre los métodos de clasificación y minería de datos

3.1. CLUSTER

Cluster es un conjunto de objetos que poseen características similares. El análisis cluster busca particionar un conjunto de objetos en grupos, de tal forma que los objetos de un mismo grupo sean similares y los objetos de grupos diferentes sean disímiles.

3.1.1. MEDIDAS DE SIMILARIDAD

Reconocer objetos como similares o disimiles es lo más importante para el proceso de clasificación. Cuando contamos con aspectos cuantitativos, este aspecto de similaridad se encuentra ligado al concepto de métrica o medidas de distancia y cuando contamos con aspectos cualitativos, hablamos de coeficientes de asociación, que se usan para datos en escala nominal.

Las medidas de similaridad usadas con más frecuencia son:

1. Medidas de distancia
2. Coeficientes de correlación
3. Coeficientes de asociación
4. Coeficientes de probabilidad

Medidas de distancia

Las medidas de distancia de uso más frecuente son:

1. Distancia euclidiana definida por

$$d_{ij} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2}$$

2. Distancia Manhattan o distancia por cuerdas. Se define por:

$$d_{ij} = \sum_{k=1}^p |X_{ik} - X_{jk}|$$

3. Distancia de Chebychev. Calcula la discrepancia más grande en alguna de las dimensiones

$$d_{ij} = \max_{k=1 \dots p} |X_{ik} - X_{jk}|$$

Coefficientes de correlación

Usamos el coeficiente de correlación para examinar el grado de similitud que hay entre dos variables numéricas. El coeficiente de correlación r , es un valor real entre -1 y 1 . Si $r = 1$ las variables están perfectamente correlacionadas, si $r = -1$ la correlación es negativa y si $r = 0$ entonces no hay correlación. Es decir que cuando r es positivo, las variables tienen un comportamiento similar, ambas crecen o decrecen al mismo tiempo. Y cuando r es negativo si una variable crece la otra decrece.

El más conocido es el coeficiente de correlación de Pearson, el cual determina el grado de de correlación o asociación lineal entre casos. Esta definido por:

$$r_{jk} = \frac{\sum_i (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2} \sqrt{\sum_i (X_{ik} - \bar{X}_k)^2}}$$

con

$$i = 1, \dots, p$$

Donde X_{ij} es el valor de la variable i para el caso j , y \bar{X}_j es la media de todas las variables que definen el caso j .

Coefficientes de asociación

Se usan principalmente cuando tenemos datos en escala nominal. Cada variable toma los valores de 0 (ausencia) y 1 (presencia) de un atributo.

Coefficientes de probabilidad

BUSCAR TEORÍA

3.1.2. MÉTODOS JERÁRQUICOS DE ANÁLISIS CLUSTER

Los métodos jerárquicos se dividen en aglomerativos y disociativos. Los métodos aglomerativos comienzan el análisis con tantos grupos como individuos haya. A partir de estos datos iniciales se van formando grupos, de forma ascendente, hasta que al final del proceso todos los casos tratados quedan en un mismo grupo.

Los métodos disociativos, constituyen el proceso inverso al anterior. Comienzan con un cluster que contiene todos los casos tratados y a partir de este grupo inicial, a través de varias divisiones, se van formando grupos cada vez más pequeños. Al final del proceso se tienen tantos grupos como casos.

A continuación vamos a presentar algunos métodos jerárquicos aglomerativos.

Método de la distancia mínima

En este método se considera que la distancia o similitud entre dos clusters viene dada por la mínima distancia (o máxima similitud) entre sus componentes.

Así, tras efectuar la etapa k -ésima, tenemos ya formados $n - k$ clusters, la distancia entre los clusters C_i (con n_i elementos) y C_j (con n_j elementos) sería:

$$d(C_i, C_j) = \min_{\substack{x_l \in C_i \\ x_m \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i; m = 1, \dots, n_j$$

En el nivel $k + 1$ se unirán los clusters C_i y C_j si

$$d(C_i, C_j) = \min_{\substack{i_1, j_1 = 1, \dots, n-k \\ i_1 \neq j_1}} \{d(C_{i_1}, C_{j_1})\} =$$

$$= \min_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \left\{ \min_{\substack{x_l \in C_{i_1} \\ x_m \in C_{j_1}}} \{d(x_l, x_m)\} \right\} \quad l = 1, \dots, n_{i_1}; m = 1, \dots, n_{j_1}$$

Método de la distancia máxima

En este método se considera que la distancia o similitud entre dos clusters hay que medirla atendiendo a sus elementos más dispares, es decir la distancia entre dos clusters viene dada por la máxima distancia entre sus componentes. Así, en la etapa k -ésima ya tenemos $n - k$ clusters, la distancia entre los clusters C_i y C_j (con n_i y n_j elementos respectivamente), serán:

$$d(C_i, C_j) = \max_{\substack{x_l \in C_i \\ x_m \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i; m = 1, \dots, n_j$$

En el nivel $k + 1$ se unirán los clusters C_i y C_j si

$$\begin{aligned} d(C_i, C_j) &= \min_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \{d(C_{i_1}, C_{j_1})\} = \\ &= \min_{\substack{i_1, j_1=1, \dots, n-k \\ i_1 \neq j_1}} \left\{ \max_{\substack{x_l \in C_{i_1} \\ x_m \in C_{j_1}}} \{d(x_l, x_m)\} \right\} \quad l = 1, \dots, n_{i_1}; m = 1, \dots, n_{j_1} \end{aligned}$$