

Segmentación y alertas para SARLAFT

Álvaro J. Riascos Villegas ¹ Natalia Serna Borrero ²

¹Universidad de los Andes-Quantil

²Universidad Icesi-Quantil

Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- Canales internos
- Canales externos
- Jurisdicciones
- 6 Resultados mixtura
 - Clientes
 - Productos
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales

Contenido

- 1 **Introducción**
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- 6 Resultados mixtura
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales

Introducción

En esta presentación mostramos y comparamos las dos metodologías propuestas para la segmentación del SARLAFT: k-medias y modelos de mixtura; para los cuatro factores de riesgo clientes, productos, canales de distribución, y jurisdicciones.

Contenido

- 1 Introducción
- 2 Metodología**
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- 6 Resultados mixtura
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales
 - Canales internos
 - Canales externos
 - Jurisdicciones

Metodología

- 1 Recopilar datos
- 2 Análisis de datos
- 3 Selección de variables
- 4 Aplicación de técnicas estadísticas para la segmentación
- 5 Presentación de resultados

Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables**
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- Canales internos
- Canales externos
- Jurisdicciones
- 6 Resultados mixtura
 - Clientes
 - Productos
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales

Selección de variables

Cientes

- Número de transacciones al mes: promedio a lo largo del trimestre del total de transacciones realizadas en cada mes.
- Monto: valor promedio de todas las transacciones realizadas.
- Ingresos: promedio de los ingresos totales reportados, donde los ingresos totales son la suma de los ingresos reportados y los ingresos adicionales reportados.
- Egresos: promedio de los egresos reportados.
- Patrimonio: promedio del patrimonio reportado.
- Activos: promedio de los activos reportados.
- Pasivos: promedio de los pasivos reportados.
- Actividad económica: código CIU a dos dígitos.

Selección de variables

Productos

- Número de transacciones al mes: promedio a lo largo del trimestre del total de transacciones realizadas por cada producto.
- Monto: promedio a lo largo del trimestre del valor total de las transacciones realizadas por cada producto.
- Cantidad de clientes: cantidad de clientes que realizan transacciones por cada producto.
- Naturaleza: variable categórica con dos niveles, captación y colocación.
- Nicho de mercado: variable categórica con tres niveles, microempresa, PYME, y todos.
- Característica: variable categórica con dos niveles, disponible y no disponible.

Selección de variables

Canales internos

- Número de transacciones al mes: promedio a lo largo del trimestre del total de transacciones realizadas en cada sucursal.
- Monto: promedio a lo largo del trimestre del valor total de las transacciones realizadas en cada sucursal.
- Cantidad de clientes: cantidad de clientes que realizan transacciones en cada sucursal.
- Casos: número de casos que se han presentado en cada sucursal.
- Regional: regional a la que pertenece la sucursal.

Canales externos

- Número de transacciones al mes: promedio a lo largo del trimestre del total de transacciones realizadas por cada intermediario.
- Monto: promedio a lo largo del trimestre del valor total de las transacciones realizadas por cada intermediario.
- Cantidad de clientes: cantidad de clientes que realizan transacciones por cada intermediario.
- Característica: variable categórica con dos niveles: presencial y no presencial.

Selección de variables

Jurisdicciones

- Número de transacciones al mes: promedio a lo largo del trimestre del total de transacciones realizadas en cada municipio.
- Monto: promedio a lo largo del trimestre del valor total de las transacciones realizadas en cada municipio.
- Cantidad de clientes: cantidad de clientes que realizan transacciones en cada municipio.
- Minería ilegal: indicador de si en el municipio se realizan actividades de minería ilegal.
- Drogas: indicador de si en el municipio se realizan actividades asociadas a la venta de estupefacientes o cultivos ilícitos.
- Trata de blancas: indicador de si en el municipio se realizan actividades de comercio sexual o trata de personas.

Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación**
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- 6 Resultados mixtura
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales
 - Clientes
 - Productos
 - Canales internos
 - Canales externos
 - Jurisdicciones

Comparación de técnicas

K-medias

- No permite usar variables categóricas a menos que las considere como numéricas o se realice la segmentación de forma separada por categoría.
- Ineficiente en la generación de alertas.
- Hay que asumir una distribución para las alertas que tal vez no corresponde con la forma como son generados los clusters. Testeamos si los datos en los clusters siguen una normal y vamos la probabilidad para definir las alertas. Si no se cumple normalidad usamos la distribución empírica.
- Se puede tener a un mismo individuo en varias categorías.

Modelos de Mixtura

- Permite usar variables categóricas. Se estima una distribución multinomial para esas variables y una normal multivariada para las continuas.
- Genera alertas teniendo en cuenta las variables categóricas.
- Las alertas siguen una distribución normal con la que también se modelan los clusters. Calculamos la distancia mahalanobis al centro del cluster y las alertas son los individuos con distancias atpicas.
- No se puede tener un individuo en varias categorías al mismo tiempo.

Comparación de técnicas

Redes Neuronales

- No es un algoritmo de segmentación, sino que debe ser combinado con técnicas de aprendizaje no supervisado para poder crear clusters.
- Se enfoca en una variable objetivo.
- Requiere muchos datos.

K-medias

Para crear los clusters:

- 1 Se asigna aleatoriamente un número de 1 a k a cada una de las observaciones de la base de datos.
- 2 Para cada cluster se calcula el centroide. El centroide es un vector que contiene el promedio de las p variables que entran en la segmentación.
- 3 Se asigna cada observación al cluster cuyo centroide es más cercano a sus características. En nuestro caso la cercanía se define en términos de la distancia euclidiana.
- 4 Se repiten los pasos 2 y 3 hasta que las asignaciones de las observaciones a los clusters no cambien.

K-medias

Para generar las alertas:

- Suponemos que las características de los clusters se distribuyen normal multivariadas con parámetros

$$\mu = [\mu_1 \mu_2 \cdots \mu_p] \text{ y } \Sigma = \begin{pmatrix} \sigma_1^2 & \cdots & \rho\sigma_1\sigma_p \\ \vdots & \ddots & \vdots \\ \rho\sigma_p\sigma_1 & \cdots & \sigma_n^2 \end{pmatrix}.$$

- Realizamos pruebas de normalidad para ver si el supuesto es válido (Shapiro-Wilk).
- Las alertas serán aquellos clientes para los que $Pr(X_i \geq x') \leq 0,01$

En los casos en los que no se cumpla normalidad usamos la distribución empírica de los datos en cada cluster.

Modelos de mixturas

Para crear los clusters:

- 1 Se estima una distribución normal multivariada para las variables continuas.
- 2 Se estima una distribución multinomial para las variables categóricas.
- 3 Se estima una distribución multinomial para unas variables dicotómicas latentes \bar{z}_k que indican a cuál cluster pertenece la observación. Sea z_{ik} la probabilidad de que i pertenezca al cluster k . Entonces, i se asigna al cluster k siempre que:
$$\bar{z}_{ik} > \bar{z}_{ik'}, \forall k \neq k'.$$
- 4 La verdadera distribución de los datos es la suma de (1) y (2) ponderadas por el tamaño relativo de cada cluster.
- 5 El número óptimo de clusters es aquel que minimice el Bayesian Information Criteria (BIC).

Modelos de mixturas

Para generar las alertas:

- Calculamos la distancia Mahalanobis de cada observación al centro del cluster:

$$M = \left((x_j - \bar{x}_j)^T \Sigma^{-1} (x_j - \bar{x}_j) \right)^{0,5} \quad (1)$$

donde \bar{x}_j es el vector de medias para las variables continuas y Σ^{-1} es la inversa de la matriz de varianzas y covarianzas.

- Calculamos el primer percentil de la distribución de la distancia Mahalanobis en cada cluster.
- Las alertas serán aquellos individuos cuya distancia supere ese umbral.

Modelos de mixturas

- Tanto en k-medias como en los modelos de mixturas se centran y se escalan las variables continuas antes de hacer la segmentación (Hastie et al., 2011).

Redes neuronales

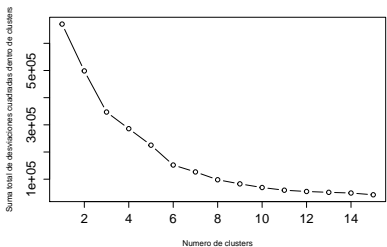
- El propósito es predecir el monto promedio por transacción usando las variables que recomienda la norma: ingresos, egresos, patrimonio, activos, pasivos, número de transacciones y actividad económica definida como el nivel de riesgo del código CIU.
- Se estimaron 30 redes neuronales y se escogió la que genera el menor RMSE.
- Las redes se estima con el algoritmo de backpropagation con funciones de activación lineales.
- Se crean clusters usando k-medias teniendo como única variable para segmentar la predicción del monto promedio por transacción que genera la red escogida.

Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias**
 - Clientes
 - Productos
- Canales internos
- Canales externos
- Jurisdicciones
- 6 Resultados mixtura
 - Clientes
 - Productos
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales

Resultados k-medias - clientes

- Para clientes usamos cuatro conjuntos de variables: monto promedio y número de transacciones; ingresos y egresos promedio; patrimonio, pasivos y activos promedio; todas las anteriores.



Resultados k-medias - clientes

Cuadro: Clusters ejercicio 1 clientes

Cluster	Monto			Transacciones			N
	Min	Max	Media	Min	Max	Media	
1	24,875,000	100,000,000	35,185,862	1.0	5.0	1.6	83
2	8,779	2,298,905	138,617	19.0	66.5	29.4	175
3	8	1,110,643	262,697	1.0	1.0	1.0	102313
4	1,049,702	3,544,439	1,865,708	1.0	5.4	1.6	12878
5	0,436	1,182,233	294,792	1.5	2.8	1.9	78798
6	4,629	5,022,190	605,765	5.5	18.4	7.8	2314
7	9,660,536	24,250,000	13,990,211	1.0	8.0	1.7	414
8	4,480	2,995,726	417,988	2.7	5.7	3.6	25136
9	0	0	0	0.0	0.0	0.0	110962
10	3,504,501	9,589,328	5,209,616	1.0	8.6	1.7	2317

Cuadro: Alertas ejercicio 1

Cluster	Identificación	Monto	Transacciones
10	1083885566	9,434,499	5.62
8	20454718	1,710,643	5.53
4	1026562240	3,471,070	4.00
4	1130600905	3,500,000	4.00
4	16733731	3,463,212	4.50
4	16369119	3,455,429	3.89
4	11801393	3,471,766	5.18

Resultados k-medias - productos

Cuadro: Clusters de productos

Cluster	Transacciones			Monto			N Clientes	Naturaleza	Nichos	Característica	N
	Min	Max	Media	Min	Max	Media					
1	3034.3	7854.3	5138.5	1,077,377,197	3,915,847,878	2,120,537,191	2560.4	1	1	1	8
2	13123.3	21148.0	17135.7	2,408,583,635	6,817,471,110	4,613,027,373	10369.5	1	1	1	2
3	34253.3	34253.3	34253.3	55,478,413,350	55,478,413,350	55,478,413,350	54564.0	0	3	0	1
4	1.0	4127.3	230.6	106,205	571,193,926	56,837,767	110.4	1	1	1	57
5	28190.0	49003.3	35320.8	4,018,767,704	16,499,981,252	9,275,136,114	19762.7	1	1	1	3
6	1.0	21094.0	5162.6	6,658	12,589,625,120	2,660,936,869	4730.6	0	3	0	5
7	1.0	319.7	54.5	1,758,653	1,477,296,805	247,474,810	43.6	1	2	1	7
8	79584.3	79584.3	79584.3	19,445,548,522	19,445,548,522	19,445,548,522	47821.0	1	1	1	1
9	3.5	694.3	151.9	59,188,151	33,627,400,998	8,500,952,638	319.0	0	3	1	8
10	6.7	237.3	122.0	4,299,410	2,136,313,976	1,070,306,693	52.5	1	3	1	2

Cuadro: Alertas de productos

Cluster	Producto	Transacc.	Monto	N Clientes	Natur.	Nichos	Caract.
3	20001	34253.3	55,478,413,350	54564	0	3	0
8	MH01	79584.3	19,445,548,522	47821	1	1	1

Resultados k-medias - canales internos

Cuadro: Clusters de canales internos

Cluster	Transacciones			Monto			N Clientes			Casos			N
	Min	Max	Media	Min	Max	Media	Min	Max	Media	Min	Max	Media	
1	34.0	124.0	87.1	438,431,668	1,058,417,733	739,078,557	1605	5731	3502.4	0	5	1.3	30
2	10.0	83.0	41.3	1,220,505,451	3,983,198,955	1,981,656,416	5916	10203	7181.4	4	10	6.0	12
3	2.0	84.0	28.3	1,385,141,772	2,362,906,189	1,887,953,771	5057	9259	7732.1	0	3	0.3	16
4	1.0	1.0	1.0	4,518,195,788	4,518,195,788	4,518,195,788	393297	393297	393297.0	12	12	12.0	1
5	105.0	142.0	130.5	2,440,098	2,133,837,294	363,268,781	1	1835	809.0	0	4	0.5	17
6	5.0	59.0	24.8	1,515,412,395	2,576,033,461	2,038,476,686	5680	7800	6839.5	27	35	31.3	4
7	14.0	120.0	64.6	618,214,835	1,846,091,082	1,367,192,835	2387	7145	5142.3	14	26	17.0	10
8	18.0	118.0	71.6	523,385,487	2,106,794,298	927,907,497	1796	6934	4361.5	5	10	7.9	17
9	95.0	95.0	95.0	59,227,893,024	59,227,893,024	59,227,893,024	119	119	119.0	0	0	0.0	1
10	9.0	117.0	60.4	729,192,664	1,759,438,610	1,181,142,477	3497	9016	5908.2	0	5	1.1	29

Cuadro: Alertas de canales internos

Cluster	Sucursal	Transacc.	Monto	N Clientes	Casos
4	1	5550	4,518,195,788	393297	12
9	95	237.6	59,227,893,024	119	0

Resultados k-medias - canales externos

Cuadro: Clusters de canales externos

Cluster	Transacciones			Monto			N Clientes			
	Min	Max	Media	Min	Max	Media	Min	Max	Media	N
1	174.7	182.3	179.1	38,585,809	44,664,887	41,395,771	145	180	160.8	4
2	213.7	277.0	236.7	61,311,395	85,314,434	72,507,455	213	272	238.8	4
3	307.7	312.7	310.2	114,099,249	126,539,919	120,319,584	250	339	294.5	2
4	241853.0	241853.0	241853.0	154,484,827,181	154,484,827,181	154,484,827,181	129765	129765	129765.0	1
5	8802.7	8802.7	8802.7	2,072,178,306	2,072,178,306	2,072,178,306	11503	11503	11503.0	1
6	59.7	79.0	71.9	11,556,381	21,518,460	17,734,702	59	84	72.3	3
7	404.7	404.7	404.7	91,937,291	91,937,291	91,937,291	400	400	400.0	1
8	247.7	292.7	274.8	95,769,322	114,369,225	102,345,802	229	317	267.7	3
9	56507.7	56507.7	56507.7	13,444,944,030	13,444,944,030	13,444,944,030	47671	47671	47671.0	1

Cuadro: Alertas de canales externos

Cluster	Canal	Transacc.	Monto	N Clientes
4	NULL	241853.0	154,484,827,181	129765
5	BALOTO	8802.7	2,072,178,306	11503
9	EFACTY	56507.7	13,444,944,030	47671

Resultados k-medias - jurisdicciones

Cuadro: Clusters de jurisdicciones

Cluster	Transacciones			Monto			N Clientes			Minería ilegal	Drogas	Otra	Trata blancas	N
	Min	Max	Media	Min	Max	Media	Min	Max	Media					
1	30093.7	30093.7	30093.7	18,990,453,728	18,990,453,728	18,990,453,728	169620	169620	169620	0.00	0.00	1.00	1.00	1
2	1.0	9919.0	2252.2	5,454,479	4,768,943,133	878,580,039	36	34485	4491	0.00	0.00	1.00	1.00	28
3	11.0	4019.3	2016.7	2,410,098	1,650,203,385	872,038,163	231	11306	4045	1.00	0.64	1.00	0.18	11
4	517.0	6961.7	3192.9	284,818,905	3,387,246,972	1,684,288,854	593	19811	8092	1.00	1.00	0.00	1.00	5
5	1816.7	6766.0	4291.3	606,753,364	2,706,345,542	1,656,549,453	5072	25179	15126	0.00	0.00	0.00	1.00	2
6	574.0	3612.7	2256.2	165,988,090	1,515,579,061	910,911,340	726	9711	4329	0.00	1.00	1.00	1.00	15
7	69266.3	69266.3	69266.3	96,800,461,965	96,800,461,965	96,800,461,965	199183	199183	199183	0.00	1.00	1.00	1.00	1
8	260.3	7705.0	2389.8	72,216,585	5,124,379,346	1,145,696,418	326	27262	5550	0.00	0.76	1.00	0.00	17
9	5771.0	9990.3	7727.7	2,440,528,372	4,809,454,979	3,611,993,459	12758	29699	22158	0.00	1.00	1.00	1.00	5
10	405.3	4791.3	2718.5	97,783,884	1,847,087,584	950,230,339	688	10434	6140	0.25	0.25	0.00	0.00	4

Cuadro: Alertas jurisdicciones

Cluster	Municipio	Transacc.	Monto	N Clientes	Minería	Drogas	Otra	Trata
1	11001	30093.7	18,990,453,728	169620	0	0	1	1
7	76001	69266.3	96,800,461,965	199183	0	1	1	1

Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- 6 Resultados mixtura
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales

Resultados mixturas - clientes

Cuadro: Clusters de clientes - Mixturas

Cluster	Ingresos		Egresos		Patrimonio		Pasivos		Activos		Transacciones		Monto		N
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	
1	0	44,500,000	0	40,764,375	(93,000,000)	200,000,000,000	0	200,000,000,000	0	370,000,000	0	66.0	0	230,880,800	3506
2	0	8,770,000	0	4,060,000	(22,119,000)	122,500,000	0	123,000,000	0	30,681,217	0	7.0	0	4,989,000	62256
3	0	38,533,332	0	22,477,056	(48,734,236)	60,701,852	0	70,550,000	0	59,510,430	0	14.3	0	3,292,305	84289
4	0	10,932,657,934	0	10,604,298,717	(24,990,000,000)	3,623,958,225	0	6,528,034,970	0	25,000,000,000	0	30.0	0	80,000,000	7564
5	0	7,269,000	0	4,316,332	(5,689,000)	8,550,000	0	9,550,000	0	7,319,537	0	6.5	0	721,620	177836

Resultados mixturas - clientes

Cuadro: Matrices de covarianza para clusters de clientes

	Ingresos	Egresos	Patrim.	Pasivos	Activos	Transacc.	Monto
C1							
Ingresos	0.0160	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Egresos	0.0000	0.0090	0.0000	0.0000	0.0000	0.0000	0.0000
Patrim.	0.0000	0.0000	95.6519	0.0000	0.0000	0.0000	0.0000
Pasivos	0.0000	0.0000	0.0000	96.1924	0.0000	0.0000	0.0000
Activos	0.0000	0.0000	0.0000	0.0000	0.1456	0.0000	0.0000
Transacc.	0.0000	0.0000	0.0000	0.0000	0.0000	25.2902	0.0000
Monto	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	66.3832
C2							
Ingresos	0.0019	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Egresos	0.0000	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
Patrim.	0.0000	0.0000	0.0019	0.0000	0.0000	0.0000	0.0000
Pasivos	0.0000	0.0000	0.0000	0.0019	0.0000	0.0000	0.0000
Activos	0.0000	0.0000	0.0000	0.0000	0.0099	0.0000	0.0000
Transacc.	0.0000	0.0000	0.0000	0.0000	0.0000	0.5920	0.0000
Monto	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2725
C3							
Ingresos	0.0354	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Egresos	0.0000	0.0216	0.0000	0.0000	0.0000	0.0000	0.0000
Patrim.	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.0000
Pasivos	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000
Activos	0.0000	0.0000	0.0000	0.0000	0.0410	0.0000	0.0000
Transacc.	0.0000	0.0000	0.0000	0.0000	0.0000	1.0216	0.0000
Monto	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1228

Resultados mixturas - clientes

Cuadro: Matrices de covarianza para clusters de clientes

	Ingresos	Egresos	Patrim.	Pasivos	Activos	Transacc.	Monto
C4							
Ingresos	41.4634	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Egresos	0.0000	42.2975	0.0000	0.0000	0.0000	0.0000	0.0000
Patrim.	0.0000	0.0000	0.3351	0.0000	0.0000	0.0000	0.0000
Pasivos	0.0000	0.0000	0.0000	0.0774	0.0000	0.0000	0.0000
Activos	0.0000	0.0000	0.0000	0.0000	41.9703	0.0000	0.0000
Transacc.	0.0000	0.0000	0.0000	0.0000	0.0000	0.7885	0.0000
Monto	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	3.0569
C5							
Ingresos	0.0019	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Egresos	0.0000	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000
Patrim.	0.0000	0.0000	0.00002	0.0000	0.0000	0.0000	0.0000
Pasivos	0.0000	0.0000	0.0000	0.00002	0.0000	0.0000	0.0000
Activos	0.0000	0.0000	0.0000	0.0000	0.0012	0.0000	0.0000
Transacc.	0.0000	0.0000	0.0000	0.0000	0.0000	0.5031	0.0000
Monto	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0112

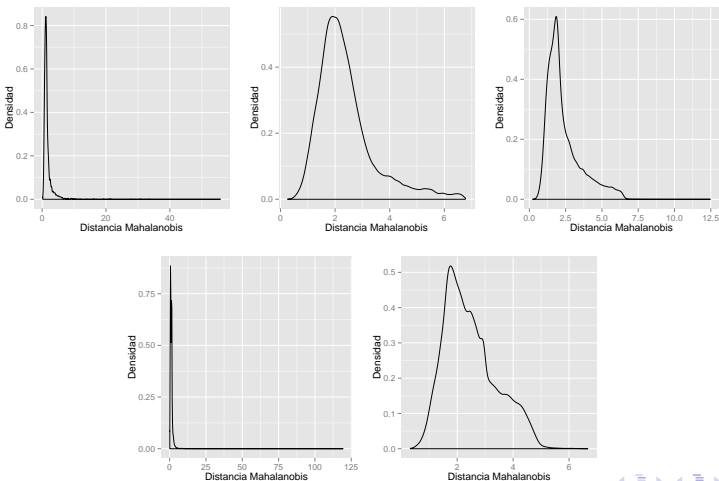
Resultados mixturas - clientes

Cuadro: Vector de probabilidades por cluster de clientes

CIU	Cluster				
	1	2	3	4	5
0	0.10381	0.07259	0.01076	0.01410	0.05897
1	0.17600	0.14102	0.10013	0.10200	0.12997
2	0.01834	0.01962	0.03593	0.03293	0.02171
3	0.00559	0.00940	0.02511	0.02642	0.01126
4	0.01156	0.01819	0.00790	0.00536	0.01762
5	0.25328	0.35149	0.70419	0.68497	0.53590
6	0.05330	0.06979	0.04429	0.06063	0.02079
7	0.16185	0.13727	0.01764	0.03125	0.02363
8	0.01908	0.01980	0.00962	0.01099	0.01609
9	0.19720	0.16084	0.04444	0.03134	0.16406

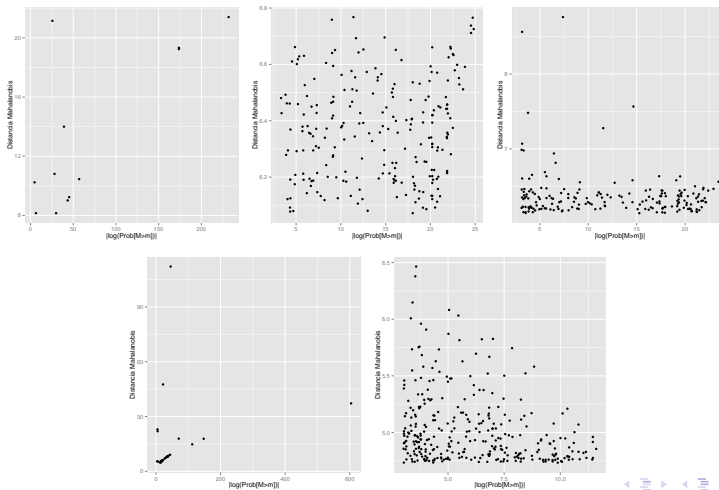
Resultados mixturas - clientes

Figura: Distribución de la distancia Mahalanobis para clientes



Clientes

Figura: Medida de atipicidad para el factor de riesgo clientes



Resultados mixturas - productos

Cuadro: Clusters de productos - Mixturas

Cluster	Transacciones			Monto			N Clientes			Colocacion	Microemp.	No disp.	N
	Min	Max	Media	Min	Max	Media	Min	Max	Media				
1	694.3	7854.3	4604.1	83,401,632	3,915,847,878	1,968,613,449	23.0	3814.0	2359.7	82 %	82 %	91 %	11
2	5.3	1392.0	401.3	38,704,344	3,340,383,288	783,564,993	3.0	1058.0	286.9	78 %	65 %	96 %	23
3	1.0	106.3	16.4	6,658	59,188,151	8,486,923	0.0	91.0	14.4	87 %	76 %	89 %	54
4	85.3	79584.3	27537.1	2,408,583,635	55,478,413,350	17,767,067,257	51.0	54564.0	20255.4	60 %	60 %	80 %	10

Cuadro: Alertas de productos para mixturas

Producto	Transacc.	Monto	NoClientes	Naturaleza	Nicho	Caract
1	694.3	3,897,327,307	1786	CAPTACION	TODOS	NODISPONIBLE
MD25	998.3	89,705,623	1058	COLOCACION	MICROEMPRESA	NODISPONIBLE
MH15	106.3	36,732,929	65	NA	NA	NA
20001	34253.3	55,478,413,350	54564	CAPTACION	TODOS	DISPONIBLE

Resultados mixturas - canales internos

Cuadro: Clusters de canales internos - Mixturas

Cluster	Transacciones			Monto			N Clientes			Casos			Regional=1	N
	Min	Max	Media	Min	Max	Media	Min	Max	Media	Min	Max	Media		
1	12.0	658.3	362.5	2,440,098	2,133,837,294	313,531,267	1	1153	585.0	0	1	0.1	0 %	12
2	237.7	5550.0	2893.8	4,518,195,788	59,227,893,024	31,873,044,406	119	393297	196708.0	0	12	6.0	50 %	2
3	2645.7	4019.3	3275.4	1,125,849,043	1,759,438,610	1,483,232,227	5057	9190	7320.0	0	6	1.6	23 %	22
4	3492.7	5169.3	4205.8	1,847,087,584	3,983,198,955	2,283,602,694	6500	10203	8113.4	0	10	2.6	36 %	14
5	631.3	1816.7	1300.5	294,858,634	923,906,932	546,386,724	744	3954	2245.0	0	17	3.4	0 %	21
6	2191.7	6295.7	3536.3	1,151,064,119	2,576,033,461	1,698,859,085	4383	7800	6148.3	6	35	18.4	63 %	16
7	1318.3	2967.0	2228.1	579,530,433	1,383,248,439	943,373,384	2387	6934	4516.3	0	15	2.9	6 %	50

Cuadro: Alertas de canales internos - mixturas

Sucursal	Transacc.	Monto	N Clientes	Casos	Regional
----------	-----------	-------	------------	-------	----------

Resultados mixturas - Canales externos

Cuadro: Clusters de canales externos - Mixturas

Cluster	Transacciones			Monto			N Clientes			Presencial	N
	Min	Max	Media	Min	Max	Media	Min	Max	Media		
1	8802.7	241853.0	102387.8	2,072,178,306	154,484,827,181	56,667,316,506	11503	129765	62979.7	67 %	3
2	59.7	404.7	219.3	11,556,381	126,539,919	67,554,757	59	400	212.2	24 %	17

Cuadro: Alertas de canales externos - mixturas

Intermediario	Transacc	Monto	N Clientes	Presencial
NULL	241853	154,484,827,181	129765	NA
CNB CAICEDONIA	404.7	91,937,291	400	0

Resultados mixturas - jurisdicciones

Cuadro: Clusters de jurisdicciones - Mixturas

Cluster	Transacciones			Monto			N Clientes			Mineria ilegal	Drogas	Otra	Trata	N
	Min	Max	Media	Min	Max	Media	Min	Max	Media					
1	30093.7	69266.3	49680.0	18,990,453,728	96,800,461,965	57,895,457,847	169620	199183	184402	0 %	50 %	100 %	100 %	2
2	1196.0	2401.7	1834.3	438,431,668	980,904,040	683,457,530	500	6037	3063	13 %	47 %	94 %	72 %	32
3	5771.0	9990.3	7593.6	2,440,528,372	5,124,379,346	3,809,069,726	12758	34485	23499	10 %	80 %	80 %	80 %	10
4	2187.0	4791.3	3233.0	553,960,208	2,355,944,540	1,344,605,287	720	14416	6913	29 %	54 %	82 %	64 %	28
5	1.0	1306.7	506.7	2,410,098	497,127,154	195,195,038	36	1597	529	24 %	47 %	88 %	47 %	17

Cuadro: Alertas de jurisdicciones - mixturas

Municipio	Transacc.	Monto	N Clientes	Mineria	Drogas	Otras	Trata
76001	69266.3	96,800,461,965	199183	0	1	1	1
23001	2238.7	980,904,040	5497	0	1	1	0
66001	9919.0	4,768,943,133	34485	0	0	1	1
66170	2187.0	553,960,208	14416	0	0	1	1
47288	930.7	294,858,634	1597	0	1	1	0

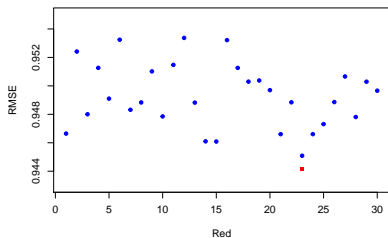
Contenido

- 1 Introducción
- 2 Metodología
- 3 Selección de variables
- 4 Técnicas de segmentación
 - K-medias
 - Modelos de mixturas
 - Redes Neuronales
- 5 Resultados k-medias
 - Clientes
 - Productos
- 6 Resultados mixtura
 - Canales internos
 - Canales externos
 - Jurisdicciones
- 7 Resultados Redes neuronales
 - Clientes
 - Productos
 - Canales internos
 - Canales externos
 - Jurisdicciones

Resultados Redes neuronales

El punto rojo es el RMSE de la red escogida.

Figura: RMSE de las distintas redes



Resultados Redes neuronales

Cuadro: Clusters a partir de predicción de la red neuronal

Cluster	Monto		Transacciones		Ingresos		Egresos		Patrimonio		Pasivos		Activos		Predicción		N
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	
1	0.01	53.47	1.00	20.20	0.00	256.89	0.00	350.00	-345.00	1599.90	0.00	892.00	5.00	1600.00	4.476	7.236	159
2	0.00	0.00	0.00	0.00	0.00	200.00	0.00	198.24	-37.66	430.00	0.00	39.17	0.00	430.00	-0.843	-0.091	8045
3	0.00	64.00	0.00	42.59	0.00	400.00	0.00	50.75	-140.00	230.00	0.00	140.00	0.00	350.00	0.619	0.882	21618
4	0.00	0.00	0.00	0.00	0.00	80.00	0.00	78.37	-64.90	520.00	0.00	65.64	0.00	520.50	-0.091	0.050	101351
5	0.00	43.00	0.00	15.00	0.00	437.03	0.00	423.12	-98.50	1486.00	0.00	1201.00	0.00	1852.67	0.240	0.432	132041
6	0.00	46.00	0.00	66.48	0.00	800.60	0.00	152.66	-1995.00	220.00	0.00	2000.00	0.00	259.00	1.380	2.404	1851
7	0.00	50.00	0.00	13.00	0.00	223.32	0.00	222.74	-300.00	349.40	0.00	500.00	6.00	350.00	2.409	4.461	438
8	0.00	0.00	0.00	0.00	0.00	94.20	0.00	87.14	-91.53	600.00	0.00	818.16	0.00	1175.96	0.050	0.240	1300
9	0.00	50.00	0.00	48.10	0.00	187.39	0.00	79.90	-145.70	1058.33	0.00	1363.59	0.00	1900.10	0.882	1.378	7211
10	0.00	100.00	0.00	27.17	0.00	84.33	0.00	76.99	-130.00	400.00	0.00	130.00	0.00	500.00	0.432	0.619	61376

Cifras de las variables continuas en millones