

Contexto Procesamiento de datos Dimensionalidad Clusterización Conclusiones

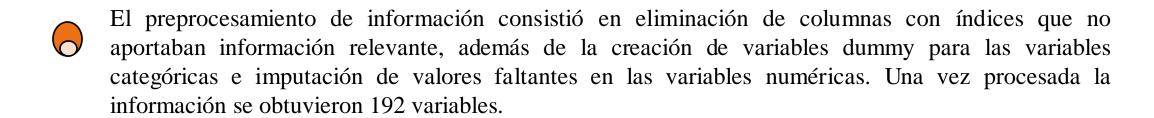
Contexto

- ¿Qué son los arquetipos de clientes? Son modelos que comparten características comunes entre los miembros que lo conforman.
- En termino generales, estas características pueden ir desde el estilo de vida hasta los comportamientos y necesidades de los clientes.
- En este caso la tarea se centra en información bancaria de 215 usuarios que se quieren clasificar de tal forma que el banco pueda entender las necesidades de cada subconjunto y planear estrategias futuras.
- Gracias a la bibliografía tanto nacional como internacional se pudo concluir que la segmentación de clientes ha sido aplicada en diversos ámbitos, tanto en el contexto nacional (Ej., de Tigo) e internacional (Rappi) y, que tiene un potencial importante de ser aplicado bajo el enfoque propuesto en este caso.

Procesamiento de datos

El dataset proporcionado por la consultora perteneciente al Banco Tibamoa (empresa ficticia), cuenta con 215 registros y 70 variables crudas.

	EDAD	ESTRATO	Probabilidad_Recomendacion	Web B Tibamoa	App B Tibamoa	Banca en linea B Tibamoa	Chat WhatsApp B Tibamoa	Telefono B Tibamoa	Cajeros electronicos	Oficinas B Tibamoa	 Banco Principal_Scotiabank Colpatria	(
0	-1.671643	-0.153944	0.668374	0.528485	0.382590	0.609123	0.544560	0.896829	0.600659	0.370000	 ()
1	-1.275213	-0.153944	0.668374	0.528485	0.382590	0.609123	0.544560	0.896829	0.600659	0.370000	 C)
2	-1.275213	-0.153944	0.320431	-2.084395	0.382590	-1.766052	0.520158	-1.199231	0.600659	0.333375	 C)
3	-1.433785	-0.153944	-2.811055	0.312743	0.101103	0.413008	0.324942	0.723760	0.405286	0.040373	 C)
4	-1.354499	-0.153944	0.668374	0.528485	0.382590	0.609123	0.544560	-1.199231	0.600659	0.370000	 C)
5 rd	ows × 192	columns										



Dimensionalidad

Debido a la cantidad de variables se buscó una reducción de la dimensionalidad aplicando estrategias como SVD y PCA. Ambos métodos sugieren 62 componentes que explican al menos el 90% de la información

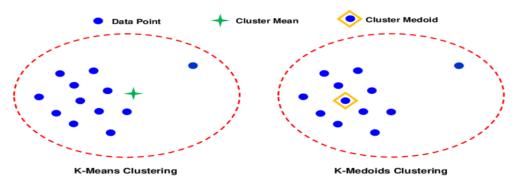
							valie	abies r	Claci	Ollauc	13 COI	1 141 3						
	N	IPS					NPS B	anca en	N	PS			NPS (Cajeros	NPS	Correo		
calificación	Probabilidad		NPS Web		NPS App		línea		whatsapp		NPSTeléfono		electrónicos		electrónico		NPS Oficinas	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%		
0	10	4,4%																
1	6	3,4%	10	4,9%	9	4,1%	10	4,4%	20	10,2%	21	9,7%	8	3,5%	14	6,7%	15	7,5%
2	2	1,2%	3	2,0%	2	1,4%	4	2,2%	9	4,7%	6	3,0%	4	2,5%	5	2,5%	10	5,6%
3	3	1,9%	5	2,6%	7	3,8%	5	2,3%	7	4,0%	6	3,5%	3	1,5%	4	2,1%	7	3,6%
4	9	4,3%	6	3,0%	6	3,3%	4	1,9%	11	4,6%	1	0,6%	2	1,0%	3	1,3%	7	3,4%
5	7	3,9%	14	6,5%	8	3,5%	9	4,4%	11	6,0%	11	4,5%	7	2,5%	7	3,5%	17	6,6%
6	4	2,3%	5	2,3%	4	2,6%	3	2,0%	6	2,0%	5	2,4%	5	2,2%	8	2,6%	10	5,2%
7	13	6,3%	11	5,3%	5	1,6%	15	7,5%	15	6,3%	16	8,1%	8	3,3%	6	3,2%	13	4,6%
8	21	8,9%	22	9,5%	28	12,0%	17	7,2%	20	8,5%	15	7,4%	23	12,7%	9	5,3%	21	9,4%
9	36	15,6%	30	14,3%	34	15,7%	30	13,9%	18	8,7%	7	4,0%	22	11,1%	11	4,4%	25	12,5%
10	104	47,7%	69	31,0%	91	40,3%	66	30,8%	59	24,2%	39	16,6%	81	33,3%	38	15,9%	75	34,7%
No conozco																		
este canal			2	1,33%	2	1,22%	5	2,40%	2	1,28%	10	5,25%	3	1,87%	13	5,54%		
No uso este																		
canal			38	17,32%	19	10,62%	47	21,12%	37	19,55%	78	35,04%	49	24,65%	97	46,99%	15	6,93%
Total																		
general	215	100%	215	100%	215	100%	215	100%	215	100%	215	100%	215	100%	215	100%	215	100%

Variables relacionadas con NDS

Para el proceso de implementación del algoritmo de clasificación se eligió PCA y con los pesos de cada componente ver cuáles eran las variables que más impactaban dentro de los primeros 10 componentes que explicaba aproximadamente el 40% de la varianza. Esto nos arrojó 11 variables todas numéricas mostradas en la tabla (además de la variable edad y estrato).

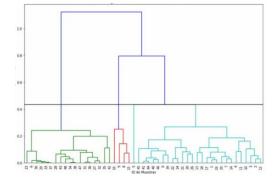
Clusterización (1)

Dado que Los algoritmos de Aprendizaje no Supervisados se utilizan para agrupar los datos no estructurados según sus similitudes y patrones distintos en el conjunto de datos, se implementaron K-medias, K-medoides y Clúster Jerárquico con el fin de revisar varios resultados y escoger el más pertinente.



K-medias: El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o clúster.

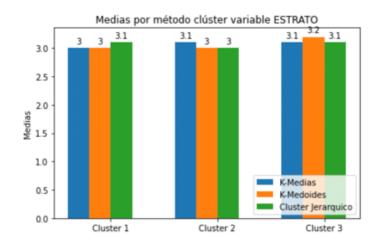
K-medoides: Divide la muestra, al igual que K-medias, en grupos minimizando la distancia entre ellos, pero asigna como centros del clúster (medoides) un punto de la muestra.

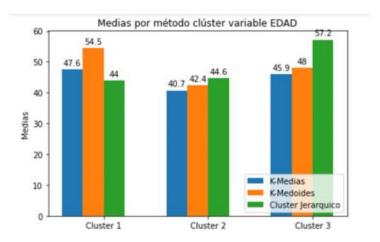


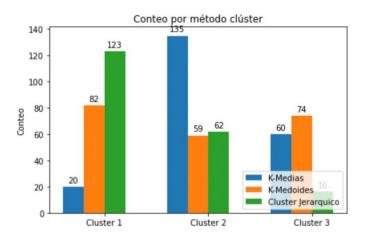
Clustering Jerárquico Aglomerativo: Se enfocan en relaciones anidadas entre los puntos de datos. El proceso inicia con cada uno de los puntos como su propio clúster y luego se van agrupando con los puntos más cercanos formando nuevos clústeres.

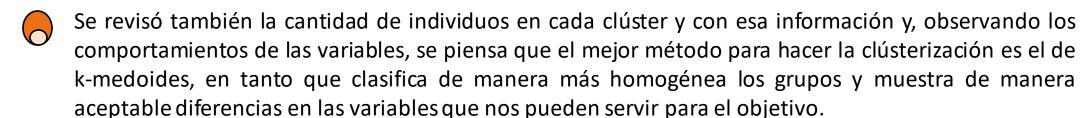
Clusterización (2)

Se implementaron los algoritmos y se comparó el comportamiento de variables que arrojaba cada uno, teniendo en cuenta cada una de las 11 variables que definimos dados los pesos encontrados en el PCA.









Clusterización (3)

Cluster	Características distintivas	Oportunidades de líneas de acción
1	 % de individuos en la muestra 38% Promedio de edad 54,5 años NPS de APP = 7 NPS chat WhatsApp= 6,5 NPS oficinas = 7,8 Probabilidad de recomendación 7,2. 	Bajo NPS o neutro, especialmente en temas relacionados con canales tecnológicos. Oportunidad de brindar atención más personalizada para que realice sus transacciones o, explicar los canales digitales. El NPS mejora para el tema de oficinas
2	 % de individuos en la muestra 27% Promedio de edad 42,4 años NPS de APP = 7,9 NPS chat WhatsApp= 6,8 NPS oficinas = 6,9 Probabilidad de recomendación 7,8. 	NPS o neutro cerca de ser promotor, valora mejor los canales tecnológicos. No valora bien oficinas presenciales. Presenta la oportunidad de desarrollar más los canales digitales para evitar que debe desplazarse a oficinas.
3	 % de individuos en la muestra 34% Promedio de edad 48 años NPS de APP = 8,9 NPS chat WhatsApp= 7,4 NPS oficinas = 7,9 Probabilidad de recomendación 9. 	NPS promotor, valora muy bien la APP y mejor el WhatsApp que el resto. También valora bien oficinas presenciales. Presenta la oportunidad de ofrecer nuevos productos y darle incentivos por recomendar nuevos clientes al banco.

Conclusiones

- Para futuros ejercicios con datos vale la pena hacer especial énfasis en la recolección de la información para evitar situaciones como la conversión de variables categóricas a binarias, que a su vez tienden a no aportar en la clusterización.
- Para nuestro caso se observa que el algoritmo k-medoides genera unos clústers más balanceados. Se presume que tiene que ver con que maneja mejor los outliers.
- Se generan 3 clústers y tiene sentido al compararlo con los análisis cualitativos realizados por la consultora donde se definieron 4 arquetipos, que, sin embargo, tuvo en cuenta más variables de comportamiento.

iiiiiMuchas Gracias!!!!