

Aprendizaje No Supervisado en la Identificación de Patrones de Desnutrición Infantil en Menores de 5 años en Medellín

Javier Ballén | Ever Contreras | Viviana Galindo | David Puerto

Resumen

La desnutrición infantil constituye un desafío global de salud pública para diversos gobiernos alrededor del mundo. Este problema es influenciado por una variedad de factores que incluyen variables de salud y nutrición, así como aspectos demográficos y socioeconómicos, que resultan en una población altamente heterogénea y que, en cualquier caso, representa un riesgo para el desarrollo y calidad de vida de los niños menores de 5 años. Con el objetivo de identificar y diferenciar los diversos perfiles de nutrición entre esta población vulnerable, se propone el uso de técnicas de aprendizaje no supervisado como Clustering combinado con PCA para la reducción de dimensionalidad, centrado en los factores más dominantes. Este enfoque permitirá descubrir y clasificar patrones distintivos de desnutrición en la ciudad de Medellín, facilitando intervenciones más dirigidas y políticas de salud pública eficientes.

1. Introducción

El problema de desnutrición infantil en Medellín afecta gravemente a los niños menores de cinco años, impactando su desarrollo físico y cognitivo. Este proyecto busca determinar: ¿Cómo las técnicas de Clustering basadas en variables de salud y nutrición pueden identificar perfiles de desnutrición en esta población? La respuesta permitiría gestionar políticas públicas más efectivas y dirigidas, optimizando la asignación de recursos.

En el contexto organizacional de la salud pública en Medellín, es crucial entender mejor las formas de desnutrición para diseñar intervenciones adecuadas. A nivel internacional, estudios como los de Striessnig y Bora (2019) en India han usado PCA y Clustering para identificar patrones de malnutrición basados en factores como el acceso a los recursos y educación. Leyso y Palatino (2020), en Filipinas, aplicaron modelos de escaneo espacial para identificar Clusters de bajo peso y sobrepeso. En Etiopía, Bitew et al. (2021) usaron modelos de Machine Learning como XGBTree y K-Nearest Neighbours para predecir factores claves de desnutrición.

Por otro lado, en Colombia, Castillo y Suarez.Ortegón (2023) investigaron la malnutrición infantil utilizando regresión logística, identificando factores demográficos importantes. Loaiza et al (2023), en Medellín, emplearon minería de datos y Machine Learning para encontrar correlaciones entre condiciones sociales y de salud.

Aun cuando varios de los estudios anteriores utilizaron desde modelos de regresión hasta Machine Learning para identificar factores de riesgo y patrones de malnutrición, las limitaciones radican en la necesidad de contar con datos precisos y actualizados. Este proyecto se centra específicamente en la aplicación de técnicas de Clustering para agrupar poblaciones según características similares de desnutrición, permitiendo una intervención más dirigida y personalizada por las autoridades de salud en Medellín.

2. Materiales y Métodos

La base *sivigila_desnutricion* se descargó de la Plataforma Nacional de Datos Abiertos de Colombia y contiene datos suministrados por la Secretaría de Salud de Colombia. La base tiene 2.802 registros de niños menores de 5 años con diagnóstico confirmado de desnutrición aguda entre los años 2016 y 2021 en Medellín.

La base tiene 26 variables, de las cuales 14 son de tipo numérico (int64), 8 de tipo carácter (string) y 4 de tipo decimal (float64). En el Anexo A se relaciona el diccionario de variables.

En el procesamiento de datos, para los valores de las variables *talla_nac*, *peso_act*, *talla_act* y *per_braqui* se ajusta el separador decimal en punto, en los casos que se requiera para que estos puedan ser interpretados correctamente por Python. Por otro lado, dado que la variable *edad* puede estar en días, meses o años, se crea una nueva variable de tipo decimal *edad_mes*, que relaciona la edad del niño en meses.

En la exploración de datos se observa que no existen valores nulos en las variables, sin embargo, se observa que las variables *peso_nac*, *per_braqui*, *talla_nac* y *edad_ges* tienen los siguientes valores en cero: 223, 853, 371 y 177, respectivamente. Estos valores no son coherentes con la naturaleza de dichas variables. Por esto, se decide imputar los valores en cero de dichas variables.

Se realizó la prueba de normalidad de Shapiro-Wilk para las variables *edad_ges* y *peso_nac*. En ambos casos el valor-p resultante es inferior a 0.05, por lo que con un 95% de confianza se rechaza la hipótesis nula de normalidad. Se resalta que para estas pruebas se excluyeron los valores en cero, dado que estos se interpretan como datos faltantes. Dado que estas variables no siguen una distribución normal, se decidió imputar los valores en cero con la mediana.

Para imputar los valores de la variable *per_braqui* se estimaron modelos de regresión XGBoost y CatBoost, empleando como variables independientes *peso_act*, *talla_act* y *edad_mes*. El modelo seleccionado para esta imputación fue el de CatBoost, que obtuvo un RMSE de 2.35, configurado con los parámetros: función de pérdida RMSE, profundidad 8 y tasa de aprendizaje 0.1.

Para imputar los valores de la variable *talla_nac*, se estimaron modelos de regresión utilizando las variables *edad_ges* y *peso_nac*. En este caso, también se eligió CatBoost, que obtuvo un RMSE de 2.37, con parámetros de función de pérdida RMSE, profundidad 5 y tasa de aprendizaje 0.1.

En el análisis descriptivo de las variables, se destaca el valor máximo del peso al nacer de 5.000 gr., que corresponde a un único individuo. Este valor está fuera de los rangos normales, pero según la literatura puede ser posible, aunque son casos muy raros. La variable *edad* tiene una dispersión alta, observando que hay niños desde los 5 días de nacido hasta los 4 años. Sin embargo, el 75% de los datos se concentran en niños de hasta 2 años (véase Anexo B).

Se destaca que hay niños con pesos actuales mayores a 20 kg., pesos que no son coherentes con los asociados a niños menores de 5 años. Por otro lado, las variables *t_lechem* y *e_complem* tiene como valores máximos 99 y 90, asumiendo que la variable está dada en meses, estos valores no pueden ser posibles, dado que los niños registrados tienen máximo 48 meses cumplidos. Adicionalmente, se recomienda dar leche materna hasta los 2 años, sin embargo, teniendo en cuenta que puede haber madres que continúan amamantando a sus hijos después de cumplida esta edad y teniendo en cuenta que los casos registrados en la base son una pequeña proporción, no se considerarán como outliers (véase Anexo C).

Dado que en el análisis de componentes principales (PCA) se requiere reducir la dimensionalidad de la base de datos, pero conservando la mayor varianza posible, esta técnica es altamente sensible a los outliers. Para evitar que los resultados del PCA se vean influenciados por valores extremos, se decide eliminar dichos valores de las variables *peso_act*, *t_lechem* y *e_complem*.

En la base de datos, el 59.6% de los registros pertenecen a niños varones, la mayoría de los niños y las niñas no están hospitalizados y cuentan con esquema de vacunación. Adicionalmente, el 58% de los individuos se encuentran en régimen contributivo, seguido de aproximadamente 39% en régimen subsidiado.

En el mapa de calor (véase Anexo D) se observa que hay una correlación positiva fuerte entre las variables edad, peso y talla actual. Esto implica que, a mayor edad del niño, se incrementa la estatura y el peso, lo cual es coherente con el desarrollo normal de un niño. Dada esta correlación, es importante que las políticas que se desarrollen entorno al crecimiento saludable del niño tengan un enfoque que incluyan el peso y la talla.

Por otro lado, existe una correlación positiva moderada entre las variables semanas de gestación, peso y talla al nacer. Lo cual es congruente con la gestación de un bebé, pues se espera que, a más semanas de gestación, el bebé pese y mida más. Además, esto influye en el riesgo de desnutrición, pues los bebés prematuros tienen más riesgo de presentar desnutrición.

3. Resultados y Discusión

La implementación del algoritmo se desarrolló en dos fases: la primera en la aplicación de PCA para determinar la influencia de las variables en la información del modelo y la reducción de su número para el paso posterior.

3.1. Primera fase (PCA):

Para el desarrollo de la primera parte, se cargan las bibliotecas requeridas para el ejercicio: en este caso el módulo de PCA de Sklearn y PCA de PyPI, OneHotEncoder y StandardScaler. En una primera aproximación se contemplaron 9 variables numéricas (*peso_nac*, *talla_nac*, *edad_ges*, *t_lechem*, *e_complem*, *peso_act*, *talla_act*, *per_braqu*, *edad_mes*) las cuales fueron estandarizadas con StandardScaler y 5 variables categóricas (*sexo*, *comuna*, *tipo_ss*, *pac_hos*, *carne_vac*), las cuales una vez fueron procesadas con OneHotEncoder, pasaron a 40 variables dummy. La dimensión de la matriz de datos para el análisis quedó en 2780 x 49.

Se aplica PCA con el módulo de Sklearn para determinar varianza explicada y varianza explicada acumulada. Para determinar el número de componentes principales a seleccionar, se grafican la varianza explicada y la varianza en función de los componentes principales (véase Anexo E) y se aplica el criterio de Kaiser para tener otra comparación. Por criterio de Kaiser se tendrían los 3 primeros componentes principales, sin embargo, estos componentes no alcanzan al 60% de la varianza explicada acumulada. Se opta entonces, por tomar los primeros 12 componentes principales, los cuales explican el 91,02% de la varianza acumulada.

Para el análisis de la influencia de las variables en los componentes principales se grafica un biplot con el módulo PCA (véase Anexo F). Se observa una gran influencia de las variables relativas a la condición física actual del niño, así como sus antecedentes alimenticios (*peso_act*, *talla_act*, *per_braqu*, *edad_mes*, *t_lechem*, *e_complem*) en el primer componente principal. Estas variables no presentan mayor influencia en el componente principal 2, el cual está formado principalmente por las variables relativas a la condición del niño al nacer (*peso_nac*, *talla_nac*, *edad_ges*). Finalmente, vale la pena resaltar que las variables categóricas seleccionadas para el análisis tienen un aporte muy pobre para ambos componentes principales analizados.

Tomando en consideración la conclusión anterior, se decidió aplicar PCA a un nuevo conjunto de datos, que solo analizaría las 9 variables numéricas antes mencionadas. Siguiendo el procedimiento arriba mencionado, por el método gráfico de varianza explicada y varianza acumulada explicada (Anexo G), se determinaron los primeros 5 componentes principales, para lograr un nivel de varianza explicada acumulada equivalente al del ejercicio previo. En el Anexo H se observan las mismas variables descritas anteriormente aportando a los componentes principales 1 y 2, pero en esta ocasión dichos componentes tienen más peso en la varianza explicada acumulada.

3.2. Segunda fase (Clustering):

Una vez se tiene el dataframe con los primeros 5 componentes principales, se avanza con la implementación de dos algoritmos de Clustering, K-Medias y DBSCAN. Así como en el caso anterior, el proceso se inicia cargando las bibliotecas requeridas: de Sklearn se incluyen KMeans y Silhouette_score y para DBSCAN de Sklearn los módulos DBSCAN y NearestNeighbors y de Kneed KneeLocator.

A continuación, se avanza en la implementación de K-medias, estimando inicialmente el número de clústeres a partir de la gráfica intracluster y el índice de Silhouette, incluidas en el Anexo I. Considerando ambas gráficas se evalúa la opción de 3 clústeres. En el Anexo J se presenta la gráfica de los datos en el plano cartesiano generado por los dos primeros componentes principales, identificando con colores a cuál de los 3 clústeres pertenecen.

Posteriormente se implementa DBSCAN, inicialmente determinando el valor óptimo del parámetro pse con la función NearestNeighbors y aplicando el diagrama para determinar el codo de la curva, se obtiene la gráfica observada en el Anexo K, la cual muestra un valor de 1 como el valor de eps óptimo. Aplicando kneed para determinar el punto exacto se obtiene un valor de 0,9262. Para el otro parámetro del algoritmo, min_samples, se evalúan valores desde 4 hasta 10 ($2 \times N^\circ$ variables) y se opta por dejar documentado el resultado con min_samples igual a 8. El resultado presentado en el mismo formato de los clústeres obtenidos por K-medias, se presenta en el Anexo L.

Después de analizar los resultados, se considera que para este caso de estudio en particular, el clustering basado en el algoritmo de K-medias ofrece el mejor desempeño, puesto que permite diferenciar claramente grupos haciendo distinciones claras entre un clúster caracterizado por mejores características físicas actuales y una mejor situación de alimentación en su primera infancia y con condiciones al nacer promedio; otro grupo con condiciones promedio al momento de nacer pero con medidas de desarrollo corporal menores, así como sus condiciones de alimentación temprana y un tercer grupo, con registros más dispersos, en general con bajas condiciones al nacer e igualmente con un desarrollo físico actual y condiciones de alimentación temprana que abarca un rango limitado hasta condiciones estándar (considerar lo observado en el Anexo J). De esta clasificación se aprecia que condiciones de bajo peso, baja talla al nacer y partos antes del tiempo de gestación óptimo son aspectos comunes en el diagnóstico de desnutrición, para un amplio espectro de condiciones de primera alimentación y condición física actual del niño, lo que indica lo susceptibles que son los niños que presentan las mencionadas condiciones en su nacimiento a desarrollar problemas de desnutrición.

Un primer aspecto por considerar en los resultados del procedimiento descrito es que a partir del análisis de PCA se aprecia que, con este set de datos, se puede eliminar del análisis la información aportada por las variables categóricas. Lo anterior muy seguramente se debe a que la base de datos corresponde completamente a niños diagnosticados con desnutrición y los datos demográficos representados en la muestra no ofrecen un nivel de varianza que aporte al modelo.

Al comparar los resultados obtenidos por los algoritmos de clustering empleados en este análisis, en el caso de DBSCAN (Anexo L) se aprecia un grupo principal el cual tiene un gran número de las muestras y otro grupo muy reducido bajo el límite inferior del grupo principal. El desempeño con DBSCAN es limitado frente a K-medias en este ejercicio debido a que la alta densidad de los datos llevó al algoritmo a clasificar esa gran masa de datos en un solo grupo, a diferencia de K-medias que se basa en la distancia a los centroides, lo que con el número adecuado de clústeres permite generar grupos claramente definidos. Adicionalmente, no se tiene una forma en la distribución de datos que no presenta formas agudas o entrelazadas, casos difíciles de analizar con K-medias. Por el contrario, la distribución de

estos datos resulta óptima para implementación del mencionado algoritmo. Por todo lo anterior, se selecciona el modelo obtenido con el algoritmo de K-medias.

Finalmente, para lograr un análisis que pueda aportar información significativa a orientar el esfuerzo de las autoridades a distribuir recursos para atender este problema de salud a lo largo de la ciudad, se requiere una muestra de información más amplia en términos geográficos y de niños con diferentes diagnósticos respecto a la condición analizada.

4. Conclusiones

Una vez realizado todo el proceso de entendimiento del problema, formulación de la pregunta, transformación y prueba de los datos, se procedió a realizar el análisis exploratorio preliminar de los datos. Teniendo esto, se realizó el cálculo de la regresión lineal (XGBoost, Catboost y el análisis exploratorio de estos resultados), obtenidos los resultados procedemos a implementar PCA con las bases de datos obtenidas, se realizan análisis con variables categoricas (tipo dummie) y sin estas, se realizan pruebas en donde podemos concluir que estas variables no aportan al modelo, por tal razón las omitimos y procedemos a correr el modelo y dar solución a la respuesta planteada previamente, en líneas generales podemos concluir que:

- El manejo los parámetros e hiperparámetros de los diversos modelos no tiene una manera comprobada que nos indique como debemos incluirlos para obtener buenos resultados, la implementación de buenos hiperparámetros se consigue con la experimentación, conocimiento y entendimiento de los datos del negocio o problema.
- Es importante considerar la cantidad de datos a procesar, ya que esto influye directamente en el costo computacional de los algoritmos utilizados. Por lo tanto, es necesario equilibrar el costo computacional con el rendimiento y la precisión del modelo, de manera que se puedan obtener resultados rápidos, confiables y que conduzcan a una solución coherente.
- Las mayores ventajas que brinda el aprendizaje no supervisado, es la facilitación del agrupamiento de los datos, este fue parte fundamental para dar solución a la pregunta que planteada para el desarrollo del proyecto.
- Al realizar proyectos de análisis de datos, es recomendable abordar la solución desde diversas perspectivas para corroborar resultados y seleccionar la opción más eficiente en términos de recursos. Además, la solución o el enfoque inicial pueden cambiar a medida que avanza el proyecto, ya que es común enfrentarse a desafíos como la calidad de los datos, la correcta aplicación de los modelos, el costo computacional y la interpretabilidad de los resultados, entre otros factores.
- Para realizar futuros análisis se sugiere utilizar una muestra de información más amplia, incluyendo información geográfica y niños con diferentes diagnósticos, no necesariamente desnutrición.
- Al concluir el ejercicio, se observa que el clustering con K-medias tiene el mejor desempeño, se logran observar tres grupos: uno con mejores características físicas actuales y buena alimentación, pero con condiciones al nacer estándar; otro con condiciones estándar al nacer, pero menor desarrollo corporal y condiciones alimentarias; y un tercero más disperso, con bajas condiciones al nacer y un amplio espectro de desarrollo físico y condiciones alimentarias, desde rangos deficientes hasta condiciones un poco más allá del estándar.

5. Bibliografía

- Ankalaki, S., G. Biradar, V., Naik P., K. K. y S. Hukkeri, G. (2024). A deep learning approach for malnutrition detection. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(06), pp. 116–138. <https://doi.org/10.3991/ijoe.v20i06.46919>
- Bitew, F.H., Sparks, C.S. y Nyarko, S.H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, 25(2), pp. 269-280. doi:10.1017/S1368980021004262
- Castillo, A.N. y Suarez-Ortegón, M.F. (2023). Dual burden of individual malnutrition in children 1–4 years: Findings from the Colombian nutritional health survey ENSIN 2015. *Pediatric Obesity*, 18(6):e13020. doi:10.1111/ijpo.13020
- Leyso, NLC. y Palatino, MC. (2020) Detecting local clusters of under-5 malnutrition in the province of Marinduque, Philippines using spatial scan statistic. *Nutrition and Metabolic Insights*, 13, pp.1-6. doi:10.1177/1178638820940670
- Loaiza, A.A., Moreno, W.G. y Ríos, J.D. (2023). Proceso de analítica de datos aplicado a la desnutrición infantil en niños de 0 a 5 años en la ciudad de Medellín. *ITM Institución Universitaria*.
- Paul, P., Arra, B., Hakobyan, M., Hovhannisyan, M.G. y Kauhanen, J. (2021). The determinants of under-5 age children malnutrition and the differences in the distribution of stunting—A study from Armenia. *PLoS ONE*, 16(5):e0249776. <https://doi.org/10.1371/journal.pone.0249776>
- Striessnig, E. y Bora, J.K. (2019). Under-Five child growth and nutrition status: spatial clustering of Indian districts. *Vienna Institute of Demography Working Papers*, 03. <https://doi.org/10.1553/0x003cb432>

6. Anexos

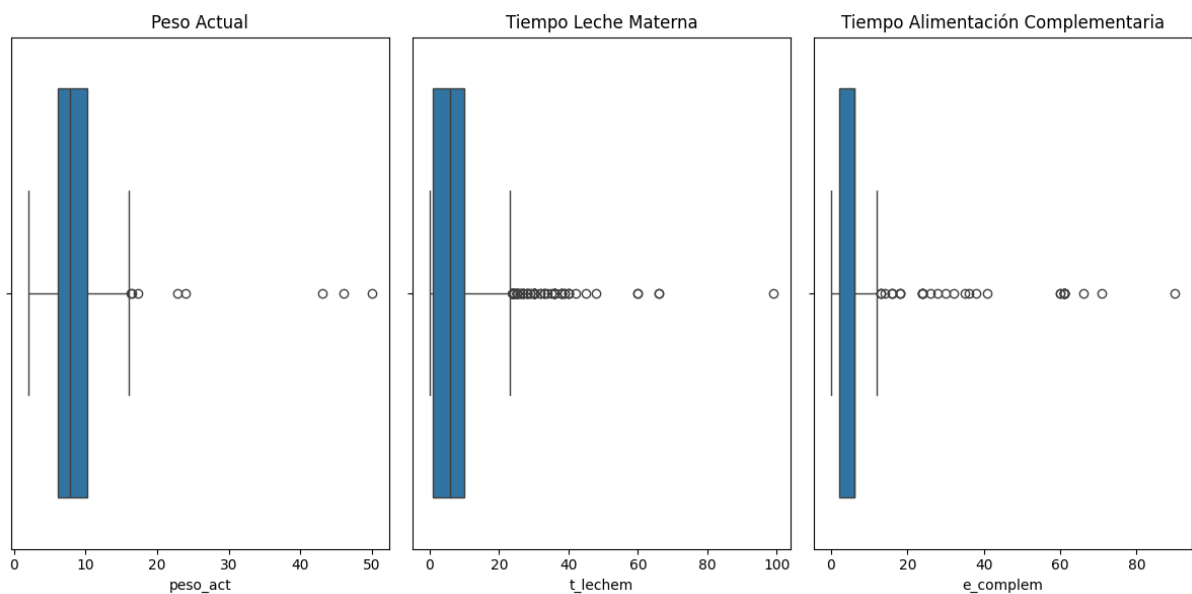
Anexo A. Diccionario de Variables.

Nombre	Tipo	Descripción Variable
ID	int64	Número consecutivo
semana	int64	semanas del año de 1 a 53
edad	int64	Edad
uni_med_	int64	Unidad de medida:\n0= No aplica, 1=Años, 2=Meses, 3=Días, 4=Horas, 5=Minutos SD=Sin informacion
sexo_	string	M=Masculino, F=Femenino, SD=Sin informacion
nombre_barrio	string	Texto asociado a la tabla de barrios definidos por la entidad territorial, Vacios se diligencian con "Sin iNformacion", Sin ubicación en zona urbana.
comuna	string	Texto asociado a la tabla de barrios definidos por la entidad territorial, Vacios se diligencian con "Sin iNformacion", Sin ubicación en zona urbana.
tipo_ss_	string	Tipo de Régimen de seguridad social:\nC= Contributivo, S=Subsidiado, P=Excepción, E=Especial,\nN= No asegurado, I= Indeterminado/Pendiente, SD=Sin informacion.
cod_ase_	string	Código de la aseguradora
fec_con_	string	Fecha de Consulta
ini_sin_	string	Fecha de inicio de síntomas
tip_cas_	int64	Tipo de caso:\n1=Sospechoso, 2= Probable , 3=Confirmado por laboratorio , 4=Confirmado por clinica , 5= Confirmado por nexa epidemiológico.
pac_hos_	int64	Paciente hospitalizado.\n1= Si, 2=No.
peso_nac	int64	Peso al nacer.\n>=900 y <=5000
talla_nac	float64	Talla al nacer.\n>=30,0 y <=55,0
edad_ges	int64	>=0 y <=45
t_lechem	int64	Tiempo de leche materna.\nnumero
e_complem	int64	Tiempo de alimentación complementaria.\nnumero
crec_dlo	int64	1= Si\n2= No
esq_vac	int64	esquema de vacunacion.\n1= Si\n2= No\n3= Desconocido
carne_vac	int64	Carnet de vacunación.\n1= Si\n2= No
peso_act	float64	peso actual
talla_act	float64	talla actual
per_braqu	float64	Perimetro braqueal.\nnumero
evento	string	Descripción del evento notificado
year_	int64	Año de notificación
edad_mes	int64	Edad en meses

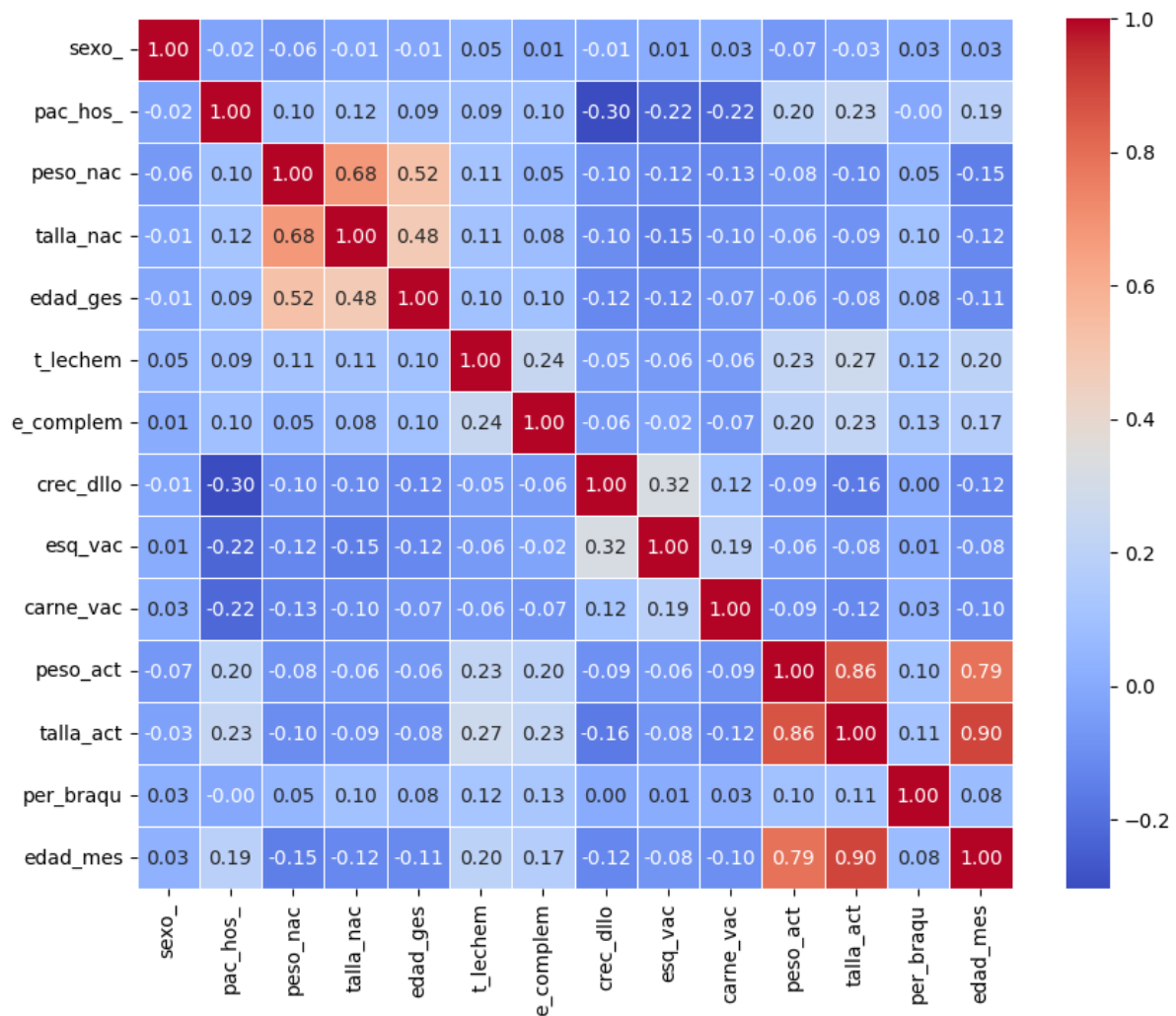
Anexo B. Estadísticas Descriptivas.

	count	unique		top	freq	mean	std	min	25%	50%	75%	max
id	2802.0	NaN		NaN	NaN	1401.5	809.012052	1.0	701.25	1401.5	2101.75	2802.0
semana	2802.0	NaN		NaN	NaN	26.838687	14.567048	1.0	14.0	27.0	39.0	53.0
sexo_	2802	2		M	1670	NaN	NaN	NaN	NaN	NaN	NaN	NaN
nombre_barrio	2802	365	SIN INFORMACION		62	NaN	NaN	NaN	NaN	NaN	NaN	NaN
comuna	2802	28	Manrique		271	NaN	NaN	NaN	NaN	NaN	NaN	NaN
tipo_ss_	2802	6	C		1620	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cod_ase_	2802	55	EPS010		754	NaN	NaN	NaN	NaN	NaN	NaN	NaN
fec_con_	2802	1286	15/06/2018		12	NaN	NaN	NaN	NaN	NaN	NaN	NaN
ini_sin_	2802	1212	01/01/1900		577	NaN	NaN	NaN	NaN	NaN	NaN	NaN
tip_cas_	2802.0	NaN		NaN	NaN	4.0	0.0	4.0	4.0	4.0	4.0	4.0
pac_hos_	2802.0	NaN		NaN	NaN	1.862241	0.344708	1.0	2.0	2.0	2.0	2.0
peso_nac	2802.0	NaN		NaN	NaN	2503.56424	927.607889	0.0	2291.0	2750.0	3050.0	5000.0
talla_nac	2802.0	NaN		NaN	NaN	41.434297	16.586429	0.0	44.0	48.0	50.0	55.0
edad_ges	2802.0	NaN		NaN	NaN	35.349036	9.535904	0.0	37.0	38.0	39.0	42.0
t_lechem	2802.0	NaN		NaN	NaN	7.02177	7.871824	0.0	1.0	6.0	10.0	99.0
e_complem	2802.0	NaN		NaN	NaN	4.821913	5.205596	0.0	2.0	6.0	6.0	90.0
crec_dilo	2802.0	NaN		NaN	NaN	1.101713	0.302325	1.0	1.0	1.0	1.0	2.0
esq_vac	2802.0	NaN		NaN	NaN	1.171663	0.481144	1.0	1.0	1.0	1.0	3.0
carne_vac	2802.0	NaN		NaN	NaN	1.437901	0.496217	1.0	1.0	1.0	2.0	2.0
peso_act	2802.0	NaN		NaN	NaN	8.28576	3.012254	2.0	6.2	7.9	10.2	50.0
talla_act	2802.0	NaN		NaN	NaN	78.168415	13.771134	45.0	67.0	76.7	89.0	116.0
per_braqu	2802.0	NaN		NaN	NaN	8.974126	6.312378	0.0	0.0	12.0	13.0	30.0
evento	2802	1	DESNUTRICION AGUDA EN MENORES DE 5 AÑOS		2802	NaN	NaN	NaN	NaN	NaN	NaN	NaN
year_	2802.0	NaN		NaN	NaN	2018.576374	1.576804	2016.0	2017.0	2018.0	2020.0	2021.0
edad_mes	2802.0	NaN		NaN	NaN	18.050471	13.677111	0.17	8.0	12.0	24.0	48.0

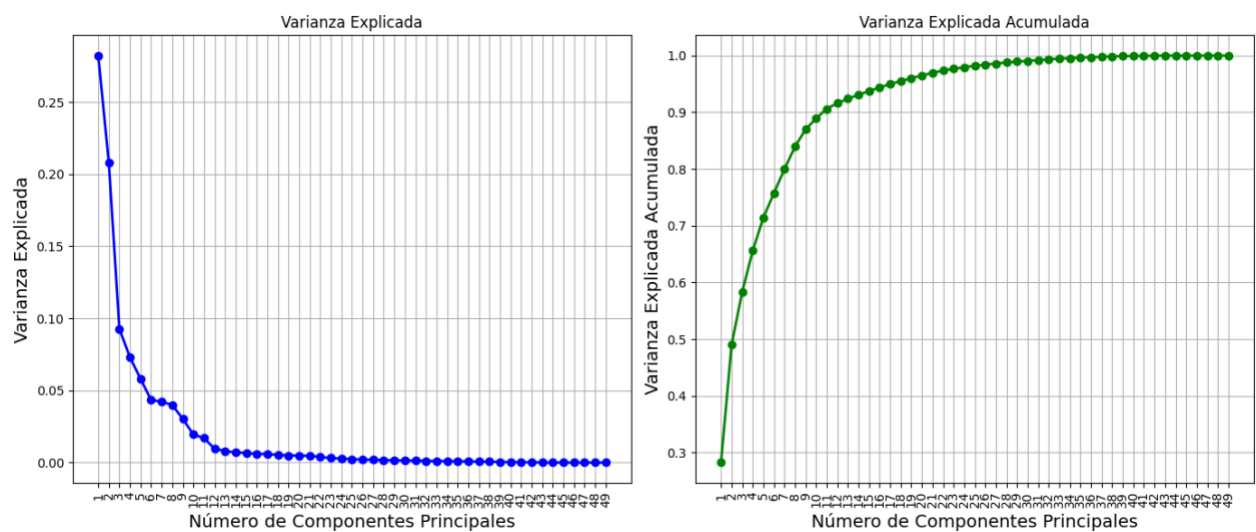
Anexo C. Boxplot de las variables: Peso Actual, Tiempo Leche Materna y Tiempo Alimentación Complementaria.



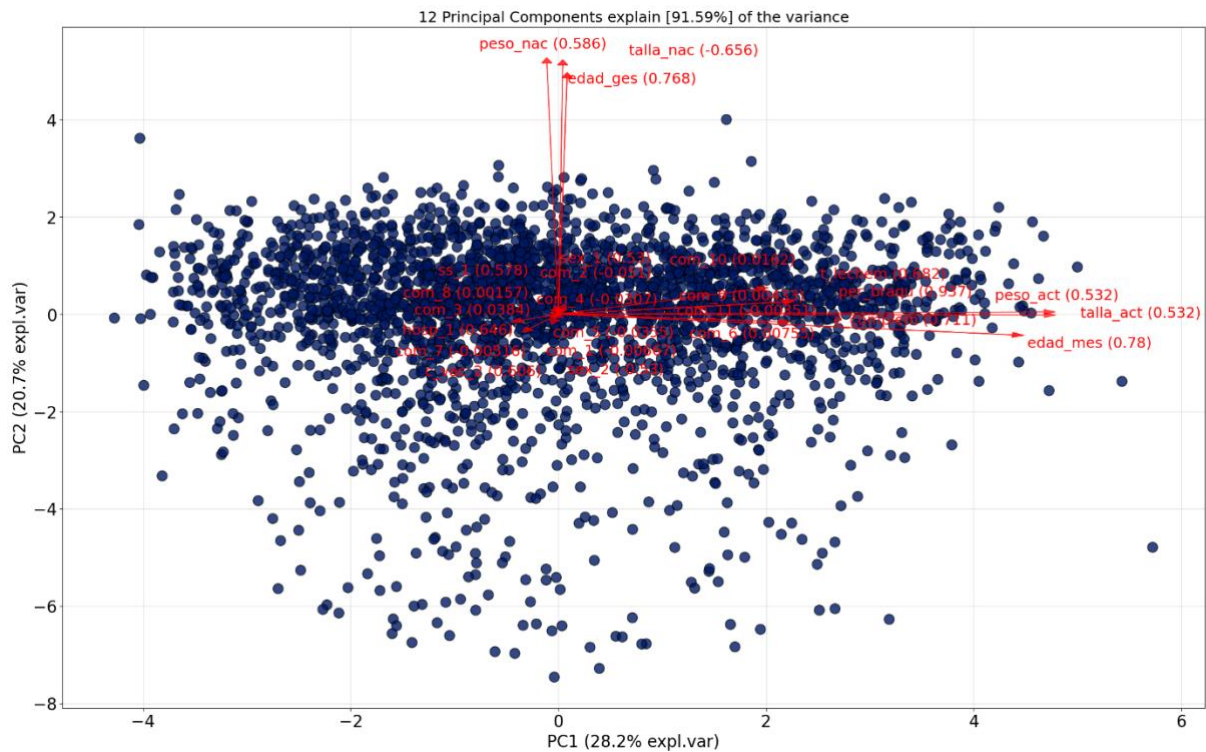
Anexo D. Matriz de Correlaciones y Mapa de Calor.



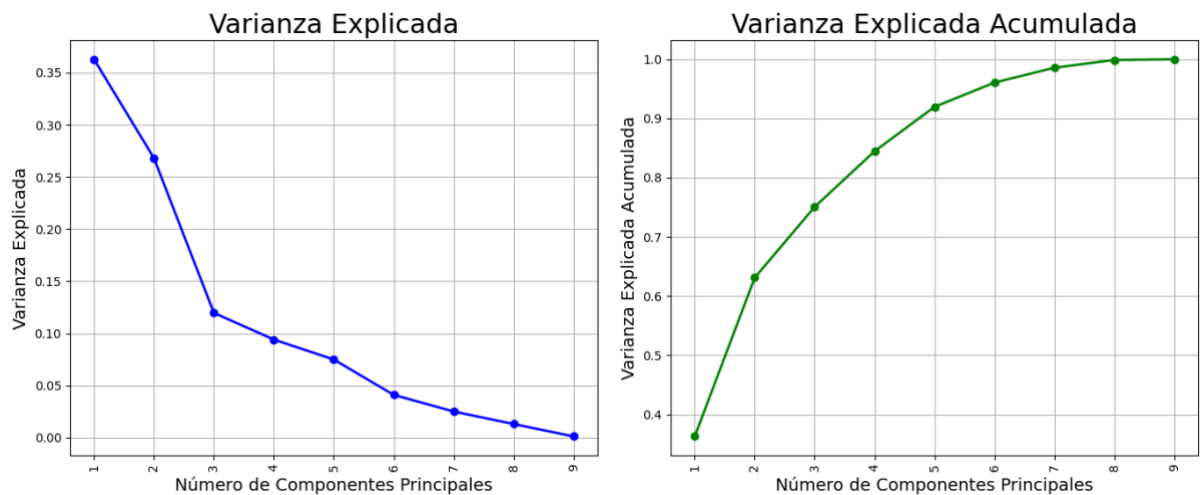
Anexo E. Varianza Explicada y Varianza Explicada acumulada – PCA para 49 variables.



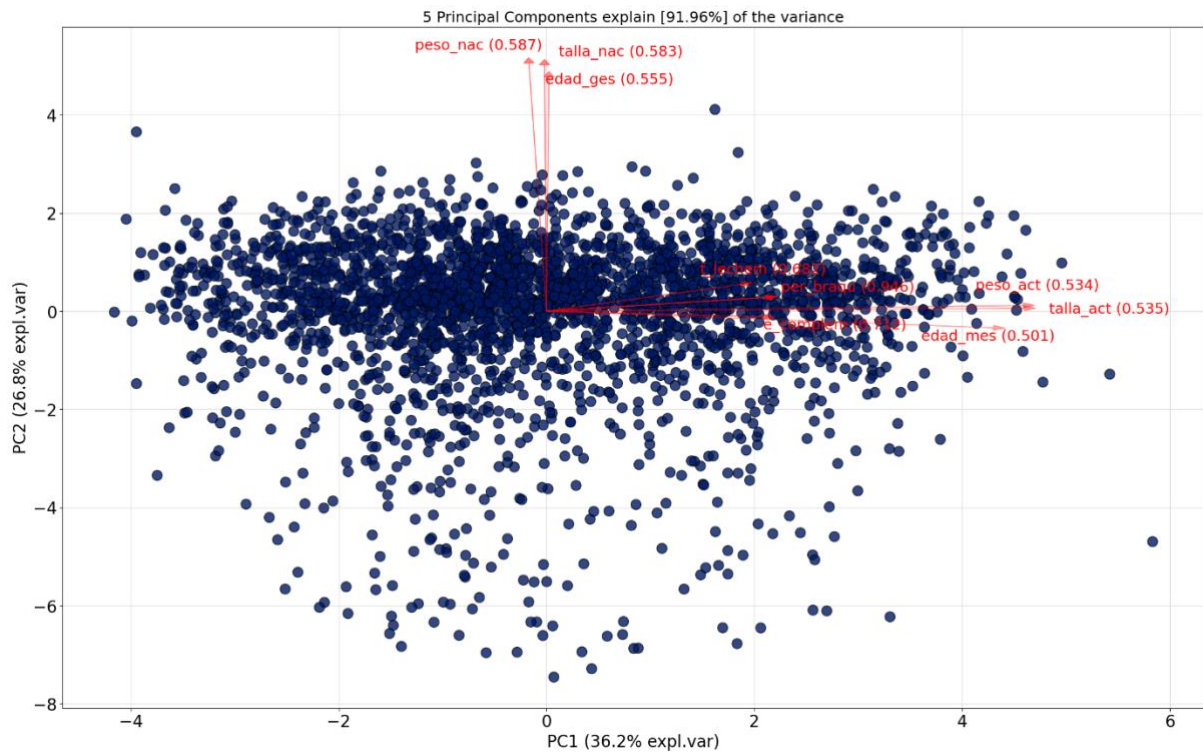
Anexo F. Biplot – PCA con 49 variables y 12 Componentes Principales.



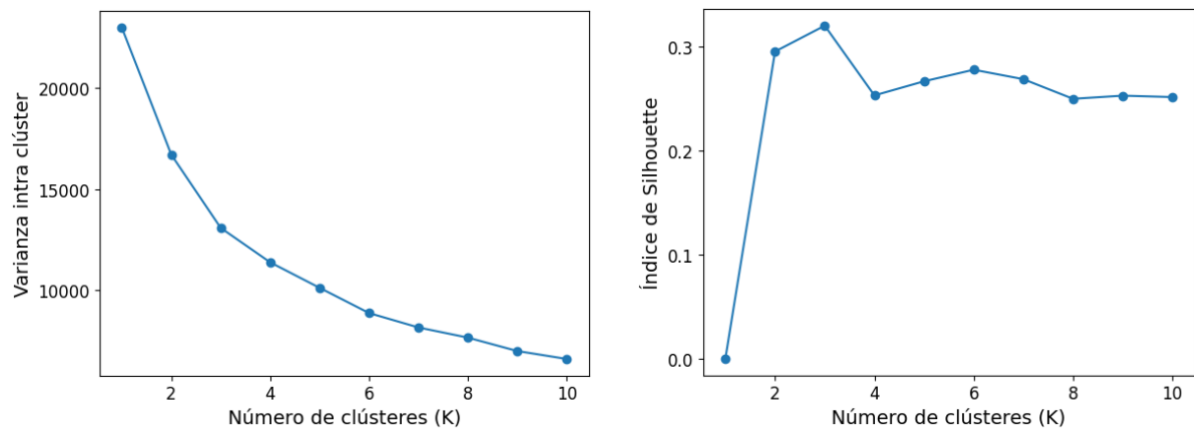
Anexo G. Varianza Explicada y Varianza Explicada acumulada – PCA para 9 variables.



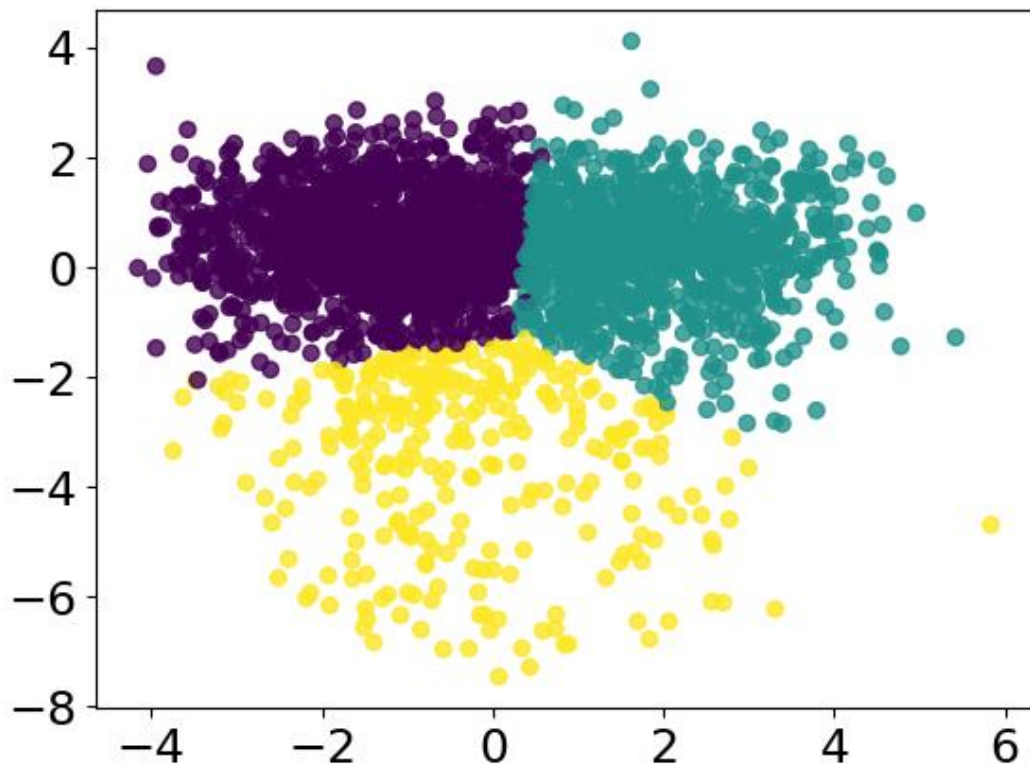
Anexo H. Biplot – PCA con 49 variables y 12 Componentes Principales.



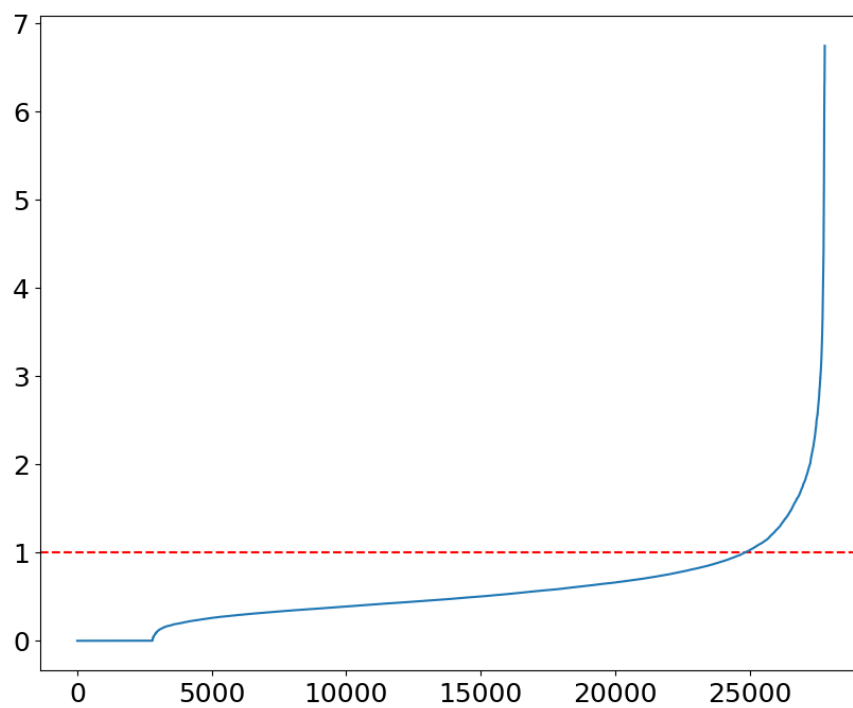
Anexo I. Número de clústeres K-medias – Varianza intra clúster e Índice de Silhouette



Anexo J. Resultado implementación K-medias con 3 clústeres.



Anexo K. Determinación del parámetro 'eps' DBSCAN.



Anexo L. Resultado implementación DBSCAN, $\text{eps} = 0,955$ y $\text{min_samples} = 8$

