

Patrones oncológicos en el Sistema Integral de Salud de Perú: Ciencia de Datos con K-Medias

Cancer Patterns in Peru's Health Insurance System: Data Science with K-Means

Wildon Rojas Paucar^{1*} <https://orcid.org/0000-0001-6590-3225>

Alberto Octavio Carranza López² <https://orcid.org/0000-0002-5147-2134>

Elena Miriam Chávez Garcés³ <https://orcid.org/0000-0002-0384-8758>

Elizabeth Luisa Medina Soto³ <https://orcid.org/0009-0005-5812-7842>

¹Universidad Nacional de Moquegua, Perú

²Universidad Nacional Mayor de San Marcos, Perú

³Universidad Nacional Jorge Basadre Grohmann Tacna, Perú

*Correo de correspondencia: wrojasp@unam.edu.pe

RESUMEN

Introducción: Las enfermedades cancerígenas se han convertido en una de las principales causas de mortalidad en América Latina. Esta investigación analiza los diferentes comportamientos asistenciales en pacientes oncológicos del Sistema Integral de Salud (SIS) en Perú, a partir de los registros de los centros hospitalarios regionales del país mediante el algoritmo K Medias.

Objetivo: Identificar patrones o agrupamientos de datos del Registro Nacional de Atenciones Oncológicas en pacientes del SIS mediante el algoritmo k-medias.

Métodos: El estudio desarrolló un enfoque de investigación no experimental, basado en la descripción y exploración de 307,145 registros de atención oncológica del Fondo Intangible Solidario de Salud (FISSAL), correspondiente al periodo 2023, donde participaron 29,962 pacientes a nivel nacional.

Resultados: Se identificaron seis agrupamientos con perfiles basados en los costos y el número de atenciones oncológicas permitidas por el SIS en las regiones afectadas del estado peruano; asimismo el desempeño del algoritmo fue escalable y eficiente en el tratamiento de los datos, al proporcionar información valiosa y significativa en la segmentación de cada agrupamiento.

Conclusión: El uso adecuado del algoritmo de K Medias permitió identificar comportamientos clave y patrones significativos, enriqueciendo la comprensión de la dinámica de los datos oncológicos del FISSAL; para la reformulación de estrategias de atención y mejores resultados en salud de los pacientes con cáncer.

Palabras clave: Patrones oncológicos; Algoritmo K-medias; Servicios de salud; Ciencia de datos

ABSTRACT

Introduction: Cancer has become one of the leading causes of mortality in Latin America. This research analyzes the varying healthcare utilization behaviors among cancer patients affiliated with Peru's Comprehensive Health Insurance System (SIS), using data obtained from regional hospital centers through the K-means algorithm.

Objective: To identify patterns or clusters within data from the National Registry of Oncology Care for SIS patients using the K-means clustering algorithm.

Methods: The study adopted a non-experimental, descriptive, and exploratory research approach, analyzing 307,145 oncology care records from the Intangible Solidarity Health Fund (FISSAL) for the year 2023, involving 29,962 patients nationwide.

Results: Six distinct clusters were identified based on profiles of cost and frequency of oncology services authorized by SIS in the affected regions of Peru. Additionally, the performance of the algorithm was scalable and efficient in processing the data, providing valuable and meaningful insights into the segmentation characteristics of each cluster.

Conclusion: Proper application of the K-means algorithm enabled the identification of key behaviors and significant patterns, enhancing the understanding of oncology data dynamics within FISSAL, thus informing the redesign of care strategies and improving health outcomes for cancer patients.

Keywords: Cancer Patterns; K-Means Algorithm; Health Services; Data Science

Recibido: 18/12/2024

Aprobado: 11/01/2025

Introducción

La enfermedad del cáncer representa un reto para todos los países en el sistema de salud pública, debido a la alta tasa de mortalidad y el costo de intervención que esta representa para su tratamiento, ^(1,2) hechos que se replican en nuestro país y

que el gobierno peruano, a través del Ministerio de Salud (MINSA) interviene mediante el Sistema Integral de Salud (SIS) garantizando a sus afiliados el tratamiento adecuado de la enfermedad, especialmente en los sectores de bajos recursos económicos. ⁽³⁻⁵⁾ El Fondo Intangible Solidario de Salud (FISSAL) es la unidad responsable de financiar dichos tratamientos a nivel nacional en todas las sedes hospitalarias del MINSA, asimismo se garantiza el anonimato y la protección de los datos personales de cada afiliado, ⁽⁶⁾ cuya información cuantitativa ha sido crucial para la presente investigación. Es importante destacar que recientemente el gobierno peruano a través de la Ley 31870 modificó la Ley Nacional del Cáncer (Ley 31336) con el objetivo de garantizar el servicio de atención hospitalaria y la cobertura universal gratuita de los servicios de salud de los pacientes oncológicos afiliados al SIS, garantizado la calidad a nivel nacional y el uso gradual de los recursos disponibles de manera responsable desde que inicia la atención oncológica hasta el final del tratamiento; ⁽⁷⁾ es importante destacar que dichas atenciones según la complejidad del caso los pacientes pueden ser derivados a otras instancias hospitalarias regionales con mayor nivel de especialización. ⁽⁸⁾

En este contexto, el FISSAL, como componente del SIS, desempeña un rol crucial en la financiación de atenciones oncológicas para los pacientes de escasos recursos. Sin embargo, la heterogeneidad en los patrones de atención, los costos asociados y la distribución geográfica de los servicios plantea desafíos importantes en la gestión eficiente de los recursos y en establecer propuestas de mejora de los resultados en salud. La falta de análisis sistemáticos con técnicas de clasificación sobre los registros de atención oncológica limita el conocimiento para identificar áreas de mejora, optimización de asignación de recursos y diseñar estrategias de intervención específicas para las necesidades de los pacientes. Por ello, el uso de técnicas en la ciencia de datos permite una mejor comprensión del comportamiento de los datos, ⁽⁹⁾ así como los algoritmos de aprendizaje no

supervisado (por ejemplo, el algoritmo K-Medias), emerge como una alternativa prometedora para identificar patrones y segmentar la información sin necesidad de estar etiquetadas, ^(10,11) por su amplia popularidad de aplicabilidad consideramos ser adecuada para el presente estudio ^(12,13) a fin de analizar los registros de atenciones oncológicas de los diferentes establecimientos hospitalarios a nivel nacional del FISSAL. Aplicar dicha puede ayudar a descubrir comportamientos en la segmentación de datos, lo cual implica una simplificación y facilitación del análisis de los datos para una mejor comprensión de los diferentes comportamientos.

En distintas investigaciones han aplicado el algoritmo K Medias con la finalidad de identificar patrones de datos e imágenes hospitalarias referentes a enfermedades oncológicas, a fin de demostrar su efectividad, por ejemplo aplicaron el algoritmo K Medias a fin de identificar regiones cancerígenas en imágenes pulmonares, en donde concluyeron que fueron muy efectivos en comparación con otros algoritmos, ⁽¹⁴⁾ asimismo otro estudio basado al cáncer de mama centro el análisis de imágenes de biopsias teñidas con Hematoxilina y Eosina (H&E), que son ampliamente utilizadas en patología para visualizar células y tejidos cuyo objetivo fue desarrollar un sistema que detecte y clasifique automáticamente el cáncer de mama a partir de imágenes histopatológicas; dentro del proceso de segmentación el algoritmo k-medias demostró ser el más efectivo en la segmentación de núcleos celulares, con una precisión del 93%, cuyo rendimiento fue superior al de otros algoritmos probados (fuzzy c-means y ACM), lo que lo convirtió en la opción preferida para la detección automática de cáncer de mama. ⁽¹⁵⁾ Otra investigación analizó los síntomas reportados por los pacientes a fin de identificar grupos de riesgo entre adultos mayores que inician tratamientos oncológicos, el uso del algoritmo k-means proporcionó una forma estructurada de análisis de datos complejos, al agrupar a los pacientes en tres clusters (baja, media y alta severidad)

conforme a 24 síntomas reportados, adicionalmente también ayudó a mejorar la atención al paciente al permitir una mejor comprensión del riesgo asociado con diferentes niveles de severidad sintomática. ⁽¹⁶⁾

Asimismo, se desarrolló una investigación consistente en analizar patrones dietéticos y su relación con el riesgo de contraer cáncer colorrectal (CRC), para ello el algoritmo k-medias fue una herramienta clave en este estudio, debido a que permitió identificar patrones dietéticos a partir de los datos recopilados, demostrando eficiencia en el procesamiento del volumen de datos, facilitando las interpretaciones de patrones específicos relacionados con los riesgos diferenciados de CRC y aportando información para estrategias preventivas basadas en la dieta, ⁽¹⁷⁾ de manera similar se desarrolló un método de selección de características para distinguir entre pacientes con cáncer colorrectal (CRC) y personas sanas utilizando un enfoque que combina el algoritmo k-means y un algoritmo de búsqueda armónica modificado con la finalidad de medir la precisión del diagnóstico temprano del cáncer colorrectal. En donde el algoritmo k-means fue esencial para el desarrollo del método propuesto en el artículo, permitiendo una identificación efectiva y eficiente de genes relevantes para el diagnóstico temprano del cáncer colorrectal adicionalmente se fusionó de manera positiva con técnicas avanzadas como la búsqueda armónica modificada demuestra su versatilidad y eficacia en aplicaciones biomédicas. ⁽¹⁸⁾

Adicionalmente, otra investigación realizó un análisis de los costos de atención médica y los factores que contribuyen a estos costos en pacientes con múltiples condiciones crónicas (MCC) en Estados Unidos, en donde el algoritmo k-medias jugó un papel crucial al permitir la identificación y análisis de agrupamientos (insuficiencia renal, trastornos respiratorios, enfermedades cardiovasculares, VIH/SIDA y trastornos mentales) de condiciones crónicas, permitiendo una

comprensión más clara y estructurada del impacto que tienen las múltiples condiciones crónicas en los costos sanitarios y la utilización de servicios médicos.

(19)

Como se ha descrito las evidencias demuestran la importancia en aplicar el algoritmo K Medias en diferentes escenarios del ámbito de la salud especialmente en analizar diferentes tipos de enfermedades oncológicas, ello implica que la cuantificación de los objetos en datos desarrolla un rol protagónico al momento de ser segmentados. A fin de ampliar el conocimiento referente a la aplicabilidad del algoritmo K Medias la presente investigación asume el desafío de la investigación concerniente a analizar el registro de atenciones hospitalarias afiliados al FISSAL del SIS, con diferentes tipos de cáncer en las diferentes regiones del país.

La presente investigación tuvo como objetivo analizar el comportamiento de los servicios de atención oncológica afiliados al SIS a nivel nacional mediante el algoritmo K Medias.

Métodos

Diseño de estudio

Se desarrolló un enfoque no experimental, siendo el diseño de la investigación descriptivo exploratorio, es decir, que no se manipularon datos, más bien se analizaron dichos datos recopilados por el FISSAL. Respecto al diseño, el algoritmo K Medias permitió identificar comportamientos y características similares de los datos de manera más precisa, ⁽²⁰⁾ ello centrado en los costos y frecuencia de atenciones oncológicas regionales.

Población y muestra

Se analizaron 307 145 registros de atención de cobertura oncológica del fondo intangible solidario de salud (FISSAL) correspondiente al año 2023. ⁽²¹⁾ La población beneficiada incluye 29 962 pacientes oncológicos registrados en el SIS, quienes recibieron atención en diversas instituciones de salud a nivel nacional. No se realizó un muestreo, ya que se trabajó con la totalidad de los registros disponibles.

Recolección de datos

La recolección de datos se obtuvo de fuentes secundarias proporcionadas por el portal del estado peruano, específicamente de FISSAL correspondiente a los registros de atenciones oncológicas. Los cuales incluyen información de costos de atención, número de atenciones y detalles de las instituciones hospitalarias a nivel nacional. Se respetaron todas las normativas de privacidad de datos vigentes.

(6)

Variables

Las variables participantes fueron el costo de atención, correspondiente al monto monetario asociado a cada atención según el tipo oncológico brindado; el número de atenciones, que definió la cantidad de atenciones recibidas por los pacientes SIS en las diferentes instituciones de salud; y finalmente, Región, que definió la ubicación geográfica del hospital o centro de atención donde se atendió al paciente.

Procedimientos

El procedimiento requirió un conjunto de datos proporcionados por FISSAL; a continuación, se desarrolló el proceso de normalización y transparencia de los datos. Seguidamente, para la manipulación de datos y descripción estadística, se utilizó R Studio por ser software libre; fue importante determinar el número ideal de agrupamientos o clústeres, para lo cual se eligió el método de la silueta; posteriormente se aplicó el algoritmo K Medias, por ser una técnica de aprendizaje no supervisado; y finalmente, la visualización de resultados se presentó mediante gráficos de variabilidad conforme a las variables involucradas.

Ética

Es importante destacar que los datos están protegidos respetando el derecho del anonimato conforme a las normas vigentes del estado peruano Ley N° 29733, ⁽⁶⁾ asimismo se garantiza su uso para fines de la presente investigación.

Resultados

En 2023, a nivel nacional se registraron 29 962 pacientes con cáncer, los cuales realizaron 307,145 atenciones según tipo de cáncer y el lugar de atención conforme al archivo de registros clínicos de las IPRESS a nivel nacional, tal como se muestra en la tabla 1 acerca de la cuantificación de los servicios de atenciones oncológicas vinculados a los siete tipos de cáncer ascienden a 307 145 (S/. 73 062 717,18) atenciones a nivel nacional, siendo el cáncer de mama con 114 760 (S/. 14 123 349,85) con mayor número de atenciones, seguido del cáncer de cuello uterino con 55 719 atenciones (S/. 8 859 520, 41), luego el cáncer de leucemia con 37 514 (S/. 28 503 669,11) atenciones, a continuación, el cáncer de linfoma con 29 811(S/. 7 604 265,44), cáncer de próstata con 27 486 (S/. 3 364 276,57), cáncer de estómago

con 21 377 (S/. 5 171 698,87), finalmente el cáncer de colon con 20 478 (S/.5 435 936,93) atenciones.

Es importante mencionar que la Leucemia a nivel nacional genera S/. 28 503 668.64 nuevos soles, servicio de mayor costo en atenciones, seguido de Cáncer de mama cuyo servicio asciende a S/.14 123 348.67 nuevos soles, Cáncer de cuello uterino S/.8 859 519.72 nuevos soles, Linfoma S/.7 604 265.05 nuevos soles, Cáncer de Colon S/.5 435 936.49 nuevos soles, Cáncer de estómago S/.5 171 698.12 nuevos soles y Cáncer de próstata con S/.3 364 276.15 nuevos soles (Tabla 1 y Figura 1).

Tabla 1. Atenciones realizadas a pacientes según tipo de cáncer

Tipo de cáncer	Número de atenciones	Costo por atención	% Costo
Cáncer de colon	20 478	5 435 936.93	7.44
Cáncer de cuello uterino	55 719	8 859 520.41	12.13
Cáncer de estomago	21 377	5 171 698.87	7.08
Cáncer de mama	114 760	14 123 349.85	19.33
Cáncer de próstata	27 486	3 364 276.57	4.60
Leucemia	37 514	28 503 669.11	39.01
Linfoma	29 811	7 604 265.44	10.41
Total	307 145	73 062 717.18	100.00

Nota: Los datos de costos de atenciones se obtuvieron del portal de datos del estado peruano ⁽²¹⁾

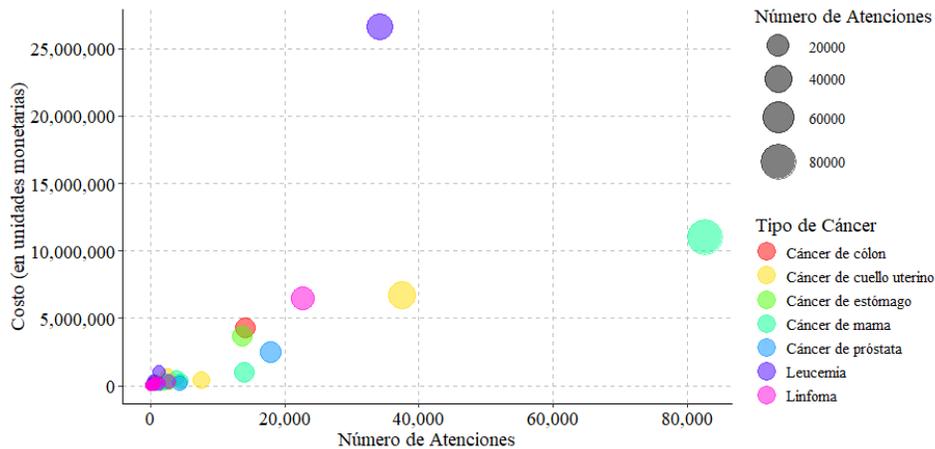


Fig. 1. Costos de atención según tipo de cáncer.

En la tabla 2 y figura 2, describe las dinámicas cuantificadas de los registros de atenciones realizadas en las diferentes regiones de Perú, podemos observar que la región Lima realizó 222 956.00 (S/. 61 169 616.15) atenciones, seguido de la región La Libertad con 34 715.00 (S/. 2 434 154.61) atenciones, seguidos de las regiones Arequipa con 12 109 (S/. 3 083 907.19), Lambayeque con 9 479 (S/. 945 983.59) atenciones, Junín con 9 133 (S/. 1 664 564.89) atenciones, Piura con 4 106 (S/. 628 335.67) atenciones, Cusco con 3 861(S/. 1 179 809.51) atenciones, Callao con 3 770 (S/. 742 633.00) atenciones, Cajamarca con 1 163 (S/. 163 093.02) atenciones, Loreto con 1 070 (S/. 223 585.60) atenciones, Ica con 896 (S/. 236 354.73) atenciones, Ancash con 837 (S/. 76 746.70) atenciones, San Martín con 678 (S/. 170 552.05) atenciones, Ayacucho con 665 (S/. 134 267.19) atenciones, Tacna con 482 (S/. 60 790.27) atenciones, Huánuco con 286 (40 455.65) atenciones, Apurímac con 258 (S/. 29 992.84) atenciones, Moquegua con 163 (S/. 20 642.72) atenciones, Tumbes con 140 (S/.13 633.94) atenciones, Ucayali con 129 (S/. 18 608.14) atenciones, Amazonas con 96 (S/. 9 949.09) atenciones, Puno con 95 (S/. 5 524.31) atenciones, Huancavelica con 33 (S/. 6 164.12) atenciones y la región Madre de Dios con 25 (S/. 3 347.89) atenciones (figura3 y figura 4).

Tabla 2. Costo de atenciones realizadas por región.

Región	Atenciones	Costo	% distribución
Amazonas	96	9 949.09	0.01
Ancash	837	76 746.72	0.11
Apurímac	258	29 992.86	0.04
Arequipa	12 109	3 083 907.47	4.22
Ayacucho	665	134 267.22	0.18
Cajamarca	1 163	163 093.17	0.22
Callao	3 770	742 633.41	1.02
Cusco	3 861	1 179 809.48	1.61
Huancavelica	33	6 164.12	0.01
Huánuco	286	40 455.64	0.06
Ica	896	236 354.78	0.32
Junín	9 133	1 664 565.16	2.28
La Libertad	34 715	2 434 155.07	3.33
Lambayeque	9 479	945 983.65	1.29
Lima	222 956	61 169 618.54	83.72
Loreto	1 070	223 585.72	0.31
Madre de Dios	25	3 347.89	0.00
Moquegua	163	20 642.72	0.03
Piura	4 106	628 335.72	0.86
Puno	95	5 524.34	0.01
San Martín	678	170 552.11	0.23
Tacna	482	60 790.21	0.08
Tumbes	140	13 633.94	0.02
Ucayali	129	18 608.15	0.03
Total	307 145	73 062 717.18	100.00

Nota: Los datos de atenciones por regiones se obtuvieron del portal de datos del estado peruano ⁽²¹⁾

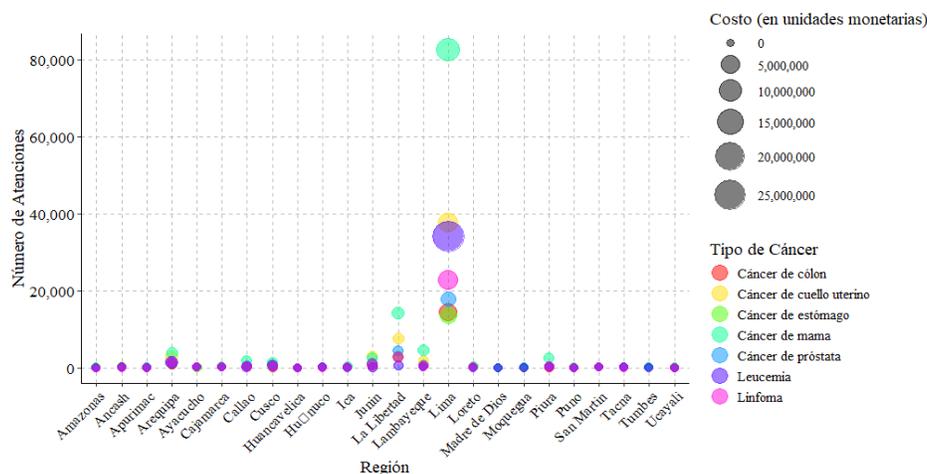


Fig. 2. Servicio de atención oncológica por regiones.

Aplicación del algoritmo K-medias

El algoritmo k-medias es una técnica que ayudará a entender la estructura de los patrones de datos, ^(22,23) por ello luego de la descripción del comportamiento general del servicio de atención oncológica en las regiones se procedió a determinar el k ideal de agrupamientos, para ello se eligió desarrollar el uso del método de la silueta, debido a su alta precisión en establecer el valor del índice, ⁽²⁴⁾ el proceso del método determinó el valor K=6 (figura 3), posteriormente se aplicó el algoritmo k-medias con la finalidad de analizar datos identificando patrones de agrupamiento de manera eficiente y automática conforme a los registros del servicio de atención oncológica.

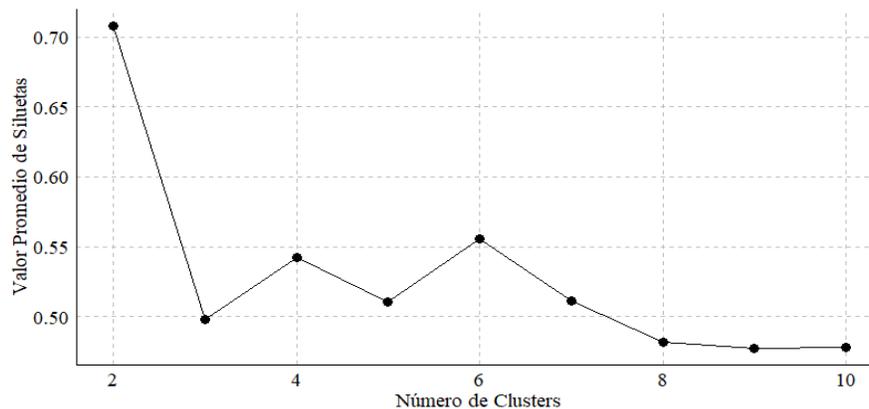


Fig. 3. Determinación del k ideal de agrupamientos.

La figura 4 muestra la descripción de los agrupamientos desarrollado por el algoritmo desde el Grupo1 hasta el Grupo 6 respectivamente, sus diferencias a

través de los colores encapsulan los tipos de enfermedades oncológicas: cáncer de colon, cáncer de cuello uterino, cáncer de estómago, cáncer de mama, cáncer de próstata, leucemia y linfoma conforme al número de atenciones y el costo que acarrea dichas enfermedades que más adelante detallaremos el valor significativo de cada agrupamiento.

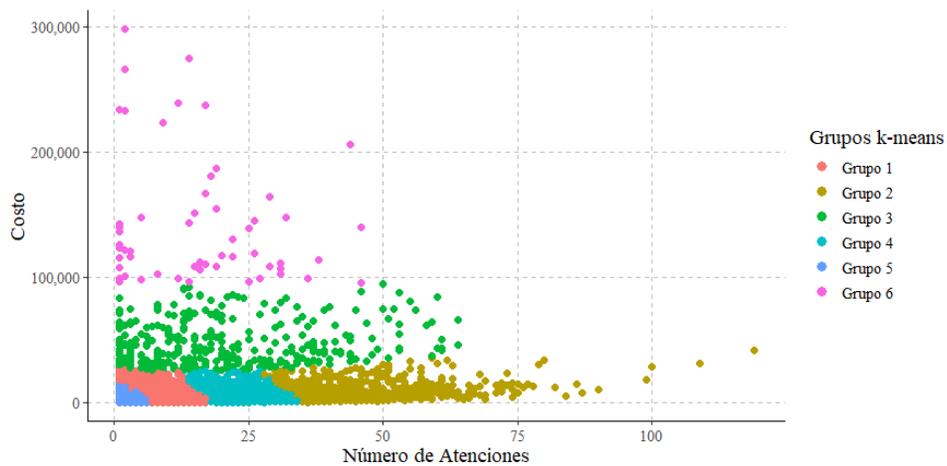


Fig. 4. Agrupamiento de los servicios de atenciones oncológica según costos.

La personalización de los agrupamientos (clústeres) en el escenario cartesiano se estableció en función de variables: número de atenciones (en el eje x) y el costo asociado al servicio de atención oncológica (en el eje y).

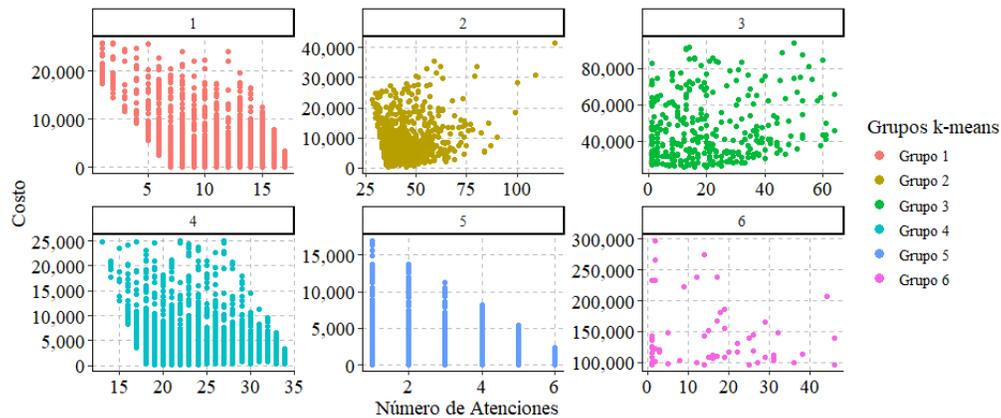


Fig. 5. Detalle de los agrupamientos.

La figura 5 apoyado por las tablas 3.1, 3.2 y 4 describe el comportamiento de cada agrupamiento conforme al plano cartesiano y que a continuación detallamos:

Agrupamiento 1:

- Se registraron entre 1 hasta 17 atenciones individuales, acumulando un total de 86 400 atenciones oncológicas con un costo de atención de S/.16 220 757.69 nuevos soles en todas las regiones, excepto Huancavelica y Ucayali,
- Se desarrolla un comportamiento elevado en las tasas de atenciones y costos, destacando las regiones de Lima, La Libertad, Arequipa y Junín.

Agrupamiento 2

- Se llevaron a cabo registros de 28 hasta 119 atenciones individuales, acumulando en el periodo 50 712 atenciones oncológicas con una inversión de S/.8 509 492.93 nuevos soles en las regiones Arequipa, Junín, La Libertad, Lambayeque y Lima.

- Las tasas de atenciones destacan un comportamiento de pocas atenciones oncológicas, pero con costos elevados, destacando las regiones Ancash, Piura, Cajamarca y Cusco.

Agrupamiento 3

- Se consignaron entre 28 hasta 64 atenciones individuales generando un total de 6 689 atenciones oncológicas, siendo el costo de inversión de S/.15 119 000.49 nuevos soles en las regiones Arequipa, Callao, Cusco, Junín, La Libertad, Lima y Loreto.
- El comportamiento en las regiones destaca un bajo número de atenciones oncológicas y costos moderados debido a que no son tan elevados, pero tampoco muy menores como los anteriores grupos.

Agrupamiento 4

- Se registraron entre 13 hasta 34 atenciones individuales, acumulando un total anual de 95 697 atenciones oncológicas, cuyo costo de inversión fue de S/.14 751 123.99 nuevos soles, en las regiones Apurímac, Arequipa, Cajamarca, Callao, Cusco, Huánuco, Junín, La Libertad, Lambayeque, Lima, Loreto, Moquegua y Piura.

El comportamiento destaca que el número de atenciones oncológicas son muy bajos y también sus respectivos costos.

Agrupamiento 5

- Se llevaron a cabo registros de 1 hasta 6 atenciones individuales, acopiando un total anual de 66 556 atenciones oncológicas cuyo monto de inversión

fue de S/.10 479 640.10 nuevos soles en todas las regiones del estado peruano.

- El grupo de regiones reportan atenciones y costos de atención muy bajos.

Agrupamiento 6

- Se consignaron de 1 hasta 46 atenciones individuales, acumulando en el año 50 712 atenciones oncológicas generando una inversión de S/.7 982 701.98 nuevos soles, en la región Lima.
- Se destaca un comportamiento por ser una región con muy alto costo de por atención, especialmente en las enfermedades de cáncer de colon, cáncer de estómago, Leucemia y Linfoma.

▪ **Tabla 3.1.** Registro de atenciones oncológicas por grupos

Región	Grupo 1		Grupo 2		Grupo 3	
	Nro. Atención	Costo	Nro. Atención	Costo	Nro. Atención	Costo
Amazonas	20.00	5 133.90	0.00	0.00	0.00	0.00
Ancash	32.00	4 773.44	0.00	0.00	0.00	0.00
Apurímac	48.00	1 006.63	0.00	0.00	0.00	0.00
Arequipa	6 070.00	1 695 793.20	35.00	13 190.05	67.00	135 400.99
Ayacucho	118.00	19 590.48	0.00	0.00	0.00	0.00
Cajamarca	464.00	45 741.58	0.00	0.00	0.00	0.00
Callao	1 207.00	263 418.97	0.00	0.00	13.00	29 239.92
Cusco	1 623.00	477 379.58	0.00	0.00	18.00	132 079.34
Huancavelica	0.00	0.00	0.00	0.00	0.00	0.00
Huánuco	7.00	1 372.53	0.00	0.00	0.00	0.00
Ica	256.00	60 428.43	0.00	0.00	0.00	0.00
Junín	3 123.00	623 272.82	354.00	65 747.14	32.00	27 809.26
La Libertad	10 030.00	605 191.63	6 359.00	470 230.83	27.00	77 516.67
Lambayeque	4 899.00	437 853.75	38.00	7 643.25	0.00	0.00
Lima	56 488.00	11 688 003.25	43 926.00	7 952 681.66	6 527.00	14 689 518.89
Loreto	237.00	30 203.78	0.00	0.00	5.00	27 435.42
Madre de Dios	0.00	0.00	0.00	0.00	0.00	0.00
Moquegua	46.00	5 201.22	0.00	0.00	0.00	0.00
Piura	1 450.00	209 001.81	0.00	0.00	0.00	0.00
Puno	0.00	0.00	0.00	0.00	0.00	0.00

San Martín	165.00	34 387.93	0.00	0.00	0.00	0.00
Tacna	102.00	11 862.30	0.00	0.00	0.00	0.00
Tumbes	15.00	1 140.46	0.00	0.00	0.00	0.00
Ucayali	0.00	0.00	0.00	0.00	0.00	0.00
Total general	86 400.00	16 220 757.69	50 712.00	8 509 492.93	6 689.00	15 119 000.49

Tabla 3.2. Registro de atenciones oncológicas por grupos

Región	Grupo 4		Grupo 5		Grupo 6	
	Nro. Atención	Costo	Nro. Atención	Costo	Nro. Atención	Costo
Amazonas	0.00	0.00	76.00	4 815.19	0.00	0.00
Ancash	0.00	0.00	805.00	71 973.28	0.00	0.00
Apurímac	19.00	875.90	191.00	28 110.33	0.00	0.00
Arequipa	1 758.00	533 386.96	4 179.00	706 136.27	0.00	0.00
Ayacucho	0.00	0.00	547.00	114 676.74	0.00	0.00
Cajamarca	79.00	6 270.14	620.00	111 081.45	0.00	0.00
Callao	648.00	57 321.13	1 902.00	392 653.39	0.00	0.00
Cusco	783.00	174 225.58	1 437.00	396 124.98	0.00	0.00
Huancavelica	0.00	0.00	33.00	6 164.12	0.00	0.00
Huánuco	43.00	4 271.91	236.00	34 811.20	0.00	0.00
Ica	0.00	0.00	640.00	175 926.35	0.00	0.00
Junín	3 004.00	431 285.02	2 620.00	516 450.92	0.00	0.00
La Libertad	11 887.00	849 088.74	6 412.00	432 127.20	0.00	0.00
Lambayeque	1 772.00	151 361.10	2 770.00	349 125.55	0.00	0.00
Lima	74 039.00	12 315 438.87	41 083.00	6 541 273.89	893.00	7 982 701.98
Loreto	264.00	21 496.36	564.00	144 450.16	0.00	0.00
Madre de Dios	0.00	0.00	25.00	3 347.89	0.00	0.00
Moquegua	18.00	2 336.81	99.00	13 104.69	0.00	0.00
Piura	1 383.00	203 765.47	1 273.00	215 568.44	0.00	0.00
Puno	0.00	0.00	95.00	5 524.34	0.00	0.00
San Martín	0.00	0.00	513.00	136 164.18	0.00	0.00
Tacna	0.00	0.00	380.00	48 927.91	0.00	0.00
Tumbes	0.00	0.00	125.00	12 493.48	0.00	0.00
Ucayali	0.00	0.00	129.00	18 608.15	0.00	0.00
Total general	95 697.00	14 751 123.99	66 754.00	10 479 640.10	893.00	7 982 701.98

Tabla 4. Descripción costos y atenciones oncológicas por tipo de enfermedad oncológica

Grupo	Tipo de cáncer	Nro. Atención	Costo	Grupo	Nro. Atención	Costo
1	Cáncer de colon	6 506	1 936 577.32	2	1 947	332 316.61

	Cáncer de cuello uterino	15 138	2 680 440.69		7 888	1 419 339.65
	Cáncer de estomago	6 911	1 865 348.21		2 394	366 872.41
	Cáncer de mama	30 821	3 905 274.37		20 584	2 395 640.31
	Cáncer de próstata	11 783	1 584 674.69		1 832	195 397.98
	Leucemia	6 061	2 293 517.68		11 273	2 996 077.09
	Linfoma	9 180	1 954 924.73		4 794	803 848.88
		86 400	16 220 757.69		50 712	8 509 492.93
3	Cáncer de colon	100	335 051.19	4	5 935	1 353 149.86
	Cáncer de cuello uterino	103	146 134.43		16 937	2 783 196.89
	Cáncer de estomago	87	316 483.03		6 093	1 129 443.76
	Cáncer de mama	134	303 229.71		44 123	5 431 934.43
	Cáncer de próstata	11	27 069.69		5 679	649 964.81
	Leucemia	5 732	12 361 233.35		8 320	2 006 592.70
	Linfoma	522	1 629 799.09		8 610	1 396 841.54
		6 689	15 119 000.49		95 697	14 751 123.99
5	Cáncer de colon	5 989	1 382 512.93	6	1	96 329.02
	Cáncer de cuello uterino	15 653	1 830 408.75		-	-
	Cáncer de estomago	5 876	1 381 516.85		16	112 034.61
	Cáncer de mama	19 098	2 087 271.03		-	-
	Cáncer de próstata	8 181	907 169.40		-	-
	Leucemia	5 258	1 580 549.86		870	7 265 698.43
	Linfoma	6 699	1 310 211.28		6	508 639.92
		66 754	10 479 640.10		893	7 982 701.98

Discusión

Los resultados obtenidos en la presente investigación demuestran la efectividad del algoritmo K-medias al momento de segmentar y analizar los patrones de atención oncológica en el SIS del Perú. La identificación de seis agrupamientos claramente diferenciados en función del número de atenciones y los costos asociados proporciona un valor detallado de la dinámica de los servicios oncológicos a nivel nacional. Los hallazgos no solo confirman la utilidad del algoritmo K-medias en el ámbito de la salud, sino que también resaltan su capacidad para manejar grandes volúmenes de datos, como los 307,145 registros analizados en este estudio. ^(10,11)

En referencia a los estudios previos, los resultados de esta investigación se alinean a los estudios previos de la aplicación del algoritmo K-medias en el análisis de

datos relacionados con enfermedades oncológicas. Por ejemplo, investigaciones como las de Dharmarajan A; Aswathy MA, Jagannath M demostraron la eficacia del K-medias en la segmentación de imágenes médicas para la detección de cáncer de pulmón y mama, respectivamente, ^(14,15) en ambos casos, el algoritmo mostró una alta precisión en la identificación de patrones, lo que coincide con nuestros hallazgos, donde el K-medias permitió identificar agrupamientos claros en los datos de atención oncológica. Además, estudios como el de Xu H y otros; Qarmiche N y otros, utilizaron el algoritmo para analizar patrones dietéticos y síntomas relacionados con el cáncer, respectivamente, ^(16,17) lo que refuerza su aplicabilidad en diferentes contextos de la salud.

De manera particular, nuestro estudio coincide con la investigación de Hajat C y otros, que utilizó el algoritmo K-medias para analizar los costos de atención médica en pacientes con múltiples condiciones crónica, ⁽¹⁹⁾ al igual que en nuestro caso, el algoritmo permitió identificar agrupamientos basados en costos y frecuencia de atenciones, lo que sugiere que esta técnica es especialmente útil para la gestión y planificación de recursos en sistemas de salud.

Respecto a las implicaciones de los resultados, los seis agrupamientos identificados en este estudio proporcionaron información valiosa para la toma de decisiones en el sistema de salud peruano. El Agrupamiento 1 presentó la mayor cantidad de atenciones (86 400) con un alto costo total de S/.16 220 757.69, concentrándose en Lima, La Libertad, Arequipa y Junín. Esto concuerda con un estudio consultado donde se identificaron algunas de estas regiones como los principales centros de demanda de servicios oncológicos debido a su infraestructura hospitalaria especializada y la alta incidencia de cáncer. ⁽²⁵⁾ El Agrupamiento 2, con 50 712 atenciones y S/.8 509 492.93 de inversión, reflejó un patrón de pocas atenciones, pero con costos elevados en regiones como Áncash,

Piura, Cajamarca y Cusco. Esta problemática podría explicarse a partir de la limitada infraestructura oncológica y el acceso restringido a tratamientos avanzados. ⁽²⁶⁾

El Agrupamiento 3, con 6 689 atenciones y S/.15 119 000.49 de inversión, mostró costos moderados, posiblemente asociados a la limitada disponibilidad de medicamentos y tratamientos especializados en Perú. No obstante, en Perú, EsSalud ofrece un anticuerpo monoclonal adicional (pertuzumab) en la etapa de tratamiento metastásico. Al propio tiempo, el Instituto Nacional de Enfermedades Neoplásicas (INEN), principalmente para la enfermedad metastásica, ha recurrido a tratamientos más probados durante el último año. Asimismo, se ha llegado a un acuerdo con las Fuerzas Armadas que permitirá a los pacientes recibir atención oncológica en el INEN y, por lo tanto, beneficiarse del uso de medicamentos de alto costo. ⁽²⁷⁾ El Agrupamiento 4, con 95 697 atenciones y costos bajos (S/.14 751 123.99), reflejó una menor complejidad en los tratamientos, posiblemente por la dificultad de acceso a servicios especializados y la falta de estrategias de detección temprana. ⁽²⁸⁾ El Agrupamiento 5, con 66 556 atenciones y costos bajos (S/.10 479 640.10), sugirió una subutilización de los servicios, posiblemente debido a la escasez de infraestructura y personal médico especializado. ^(26,28) Finalmente, el Agrupamiento 6, concentrado en Lima con 50 712 atenciones y S/.7 982 701.98 de inversión, presentó los costos más altos por atención, especialmente en cáncer de colon, estómago, leucemia y linfoma, debido a la concentración de hospitales especializados y el uso de terapias avanzadas. Esto podría significar una mayor complejidad en los tratamientos que va desde una mayor carga financiera para los pacientes y un mayor desafío para el sistema de salud. Además, la centralización de los servicios oncológicos en la capital puede generar desigualdades en el acceso a tratamientos de alta calidad en otras regiones del país. ⁽²⁵⁻²⁸⁾

En relación a las limitaciones, aunque el presente estudio proporciona información valiosa, es importante reconocer sus limitaciones; en primer lugar, los datos analizados se limitan a un solo año 2023, lo que no permite evaluar tendencias a largo plazo. Además, el estudio no considera variables socioeconómicas o geográficas que podrían influir en el acceso a los servicios oncológicos, como la distancia a los centros de salud o el nivel educativo de los pacientes, las futuras investigaciones podrían incorporar estas variables para obtener una comprensión más completa de los patrones de atención.

Las recomendaciones para futuras investigaciones ante el éxito del algoritmo K-medias en este estudio, se recomienda su aplicación en otros contextos de la salud pública, como la segmentación de pacientes con enfermedades crónicas o la optimización de recursos en hospitales. Además, sería beneficioso realizar estudios longitudinales que analicen la evolución de los patrones de atención oncológica a lo largo del tiempo. También se sugiere explorar la integración del K-medias con otras técnicas de inteligencia artificial, como el aprendizaje profundo, para mejorar la precisión y el alcance de los análisis.

Los resultados obtenidos permiten establecer las siguientes conclusiones:

- El algoritmo K-medias demostró ser una herramienta eficaz para identificar patrones y agrupamientos en los datos de atención oncológica, cuya aplicación permitió segmentar los registros en seis grupos claramente diferenciados, basados en el número de atenciones y los costos asociados. Esto confirma la utilidad del K-medias para el análisis de grandes volúmenes de datos en el ámbito de la salud, tal como se ha evidenciado en estudios previos. ^{(14)–(16)}
- Los agrupamientos identificados revelaron diferencias significativas en la demanda y los costos de los servicios oncológicos entre las regiones del Perú.

Por ejemplo, el Agrupamiento 1, que incluye regiones como Lima, La Libertad, Arequipa y Junín, mostró un alto número de atenciones y costos elevados, lo que sugiere una mayor demanda de servicios en estas áreas. En contraste, el Agrupamiento 5, que abarca todas las regiones, pero con un número muy bajo de atenciones, podría indicar barreras de acceso o una menor incidencia de cáncer en ciertas zonas. Estos hallazgos resaltan la necesidad de políticas públicas diferenciadas para abordar las disparidades regionales en la atención oncológica.

- Respecto a la concentración de casos complejos descritos en el Agrupamiento 6 evidencia una exclusiva concentración en la región Lima, demostrando costos extremadamente elevados por atención, especialmente en casos de cáncer de colon, estómago, leucemia y linfoma, lo cual se sugiere que Lima, como principal centro de referencia nacional, atiende casos más complejos y costosos, lo que justifica una mayor inversión en infraestructura y especialización en esta región.
- La identificación de agrupamientos con altos costos y demanda de servicios, como el Agrupamiento 1, sugiere la necesidad de reforzar los recursos médicos y financieros en estas regiones. Por otro lado, los agrupamientos con bajos niveles de atención, como el Agrupamiento 5, indican la importancia de implementar estrategias de promoción y acceso a los servicios oncológicos en zonas con menor cobertura.
- La presente investigación enriquece la comprensión de la dinámica de los servicios oncológicos en el Perú, adicionalmente contribuye al campo de la ciencia de datos aplicada a la salud. La aplicación exitosa del algoritmo K-medias en este contexto respalda su uso en futuras investigaciones relacionadas con la segmentación de pacientes, la optimización de recursos y la planificación de políticas públicas en salud.

- Finalmente, el uso del algoritmo K-medias es una herramienta poderosa para la segmentación y análisis de datos oncológicos en el sistema de salud peruano, cuyos resultados obtenidos no solo proporcionan información valiosa para la toma de decisiones en salud pública, sino que también abren nuevas líneas de investigación en el campo de la ciencia de datos aplicada a la salud.

Referencias bibliográficas

1. McGarvey N, Gitlin M, Fadli E, Chung KC. Increased healthcare costs by later stage cancer diagnosis. BMC Health Serv Res [Internet]. 2022 [citado el 13 de noviembre de 2024];22(1):1155. Disponible en: 1155 (2022). <https://doi.org/10.1186/s12913-022-08457-6>
2. Mehlis K, Witte J, Surmann B, Kudlich M, Apostolidis L, Walther J, et al. The patient-level effect of the cost of Cancer care – financial burden in German Cancer patients. BMC Cancer [Internet]. 2020 [citado el 13 de noviembre de 2024];20(1):529. Disponible en: <https://doi.org/10.1186/s12885-020-07028-4>
3. Altini M, Solinas L, Bucchi L, Gentili N, Gallegati D, Balzi W, et al. Assessment of Cancer Care Costs in Disease-Specific Cancer Care Pathways. Int J Environ Res Public Health [Internet]. 2020 [citado el 17 de octubre de 2024];17(13):1–13. Disponible en: <https://doi.org/10.3390/ijerph17134765>
4. Stanway S, Lodge M, Sullivan R, Diprose K, Young AM, Crisp N, et al. The UK's contribution to cancer control in low-income and middle-income countries. Lancet Oncol [Internet]. 2021 [citado el 13 de noviembre de 2024];22(9):e410–8. Disponible en: [https://doi.org/10.1016/s1470-2045\(21\)00380-6](https://doi.org/10.1016/s1470-2045(21)00380-6)

5. Stefan DC, Tang S. Addressing cancer care in low- to middle-income countries: a call for sustainable innovations and impactful research. BMC Cancer [Internet]. 2023 [citado el 13 de octubre de 2024];23(1):756. Disponible en: <https://doi.org/10.1186/s12885-023-11272-9>
6. Congreso de la República del Perú. Ley N.° 29733 / Ley de Protección de Datos Personales [Internet]. Normas Legales - El Peruano jul 3, 2011. Disponible en: <https://cdn.www.gob.pe/uploads/document/file/272360/Ley%20N%C2%BA%2029733.pdf.pdf?v=1618338779>
7. Congreso de la República del Perú. Ley 31870 / Ley que modifica la Ley 31336, Ley Nacional del Cáncer [Internet]. Lima; sep 12, 2023. Disponible en: <https://img.lpderecho.pe/wp-content/uploads/2023/09/Ley-31870-LPDerecho.pdf>
8. Ministerio de Salud. Resolución Ministerial N.° 325-2012-MINSA [Internet]. Lima: N.° 325-2012-MINSA; abr 24, 2012. Disponible en: <https://www.gob.pe/institucion/minsa/normas-legales/242173-325-2012-minsa>
9. Rojas-Paucar W, Carranza López AO, Chávez Garcés EM, Supo Gavancho KY, Gutierrez Perez JR. Análisis en la dinámica regional de datos del dengue en Perú mediante Componentes Principales, periodo 2000-2023. Revista Cubana de Investigaciones Biomédicas [Internet]. el 9 de diciembre de 2024 [citado el 12 de noviembre de 2024];43(0):e3627. Disponible en: <https://revibiomedica.sld.cu/index.php/ibi/article/view/3627/1692>
10. Ahmed M, Seraj R, Shamsul Islam SM. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics (Basel)

- [Internet]. el 12 de agosto de 2020 [citado el 12 de octubre de 2024];9(8):1295. Disponible en: <https://doi.org/10.3390/electronics9081295>
11. Golalipour K, Akbari E, Hamidi SS, Lee M, Enayatifar R. From clustering to clustering ensemble selection: A review. Eng Appl Artif Intell [Internet]. 2021 [citado el 12 de enero de 2025];104:104388. Disponible en: <https://doi.org/10.1016/j.engappai.2021.104388>
 12. Chaudhry M, Shafi I, Mahnoor M, Vargas DLR, Thompson EB, Ashraf I. A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. Symmetry (Basel) [Internet]. 2023 [citado el 13 de noviembre de 2024];15(9):1679. Disponible en: <https://doi.org/10.3390/sym15091679>
 13. Ikotun AM, Ezugwu AE, Abualigah L, Abuhaija B, Heming J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. Inf Sci (N Y) [Internet]. el 1 de abril de 2023 [citado el 12 de noviembre de 2024];622:178–210. Disponible en: <https://doi.org/10.1016/j.ins.2022.11.139>
 14. Dharmarajan A, T V. Performance Analysis on K-Means and Fuzzy C-Means Clustering Algorithms Using CT-DICOM Images of Lung Cancer. Journal of Advanced Research in Dynamical and Control Systems [Internet]. 2019 [citado el 13 de octubre de 2024];11(9-SPECIAL ISSUE):494–502. Disponible en: <http://dx.doi.org/10.5373/JARDCS/V11/20192597>
 15. Aswathy MA, Jagannath M. Performance Analysis of Segmentation Algorithms for the Detection of Breast Cancer. Procedia Comput Sci [Internet]. el 1 de enero de 2020 [citado el 12 de noviembre de 2024];167:666–76. Disponible en: <https://doi.org/10.1016/j.procs.2020.03.333>

16. Xu H, Mohamed M, Flannery M, Peppone L, Ramsdale E, Loh KP, et al. An Unsupervised Machine Learning Approach to Evaluating the Association of Symptom Clusters With Adverse Outcomes Among Older Adults With Advanced Cancer. JAMA Netw Open [Internet]. 2023 [citado el 13 de octubre de 2024];6(3):e234198. Disponible en: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2802736>
17. Qarmiche N, Kinany K El, Otmani N, Rhazi K El, Chaoui NEH. Cluster analysis of dietary patterns associated with colorectal cancer derived from a Moroccan case-control study. BMJ Health & Care Informatics Online [Internet]. 2023 [citado el 13 de octubre de 2024];30(1):e100710. Disponible en: <https://doi.org/10.1136/bmjhci-2022-100710>
18. Bae JH, Kim M, Lim JS, Geem ZW. Feature Selection for Colon Cancer Detection Using K-Means Clustering and Modified Harmony Search Algorithm. Mathematics [Internet]. 2021 [citado el 13 de octubre de 2024];9(5):570. Disponible en: <https://doi.org/10.3390/math9050570>
19. Hajat C, Siegal Y, Adler-Waxman A. Clustering and Healthcare Costs With Multiple Chronic Conditions in a US Study. Front Public Health [Internet]. 2021 [citado el 13 de octubre de 2024];8:1-10. Disponible en: <https://doi.org/10.3389/fpubh.2020.607528>
20. Saunders M, Lewis P, Thornhill A. Research methods for business students [Internet]. 9a ed. Thornhill, Adrian. Pearson; 2012 [citado el 13 de octubre de 2024]. 1-729 p. Disponible en: https://www.researchgate.net/publication/240218229_Research_Methods_for_Business_Students
21. Fondo Intangible Solidario de Salud. Plataforma Nacional de Datos Abiertos [Internet]. Atenciones de cobertura oncológica periodo 2023 FISSAL. 2023

- [citado el 13 de octubre de 2024]. Disponible en: <https://www.datosabiertos.gob.pe/dataset/atenciones-de-coertura-oncológica-periodos-2022-2023-fondo-intangible-solidario-de-salud>
22. Ali HH, Kadhum LE. K- Means Clustering Algorithm Applications in Data Mining and Pattern Recognition. En 2017 [citado el 11 de octubre de 2024]. Disponible en: <https://api.semanticscholar.org/CorpusID:36213323>
 23. Hossain MdZ, Akhtar MdN, Ahmad RB, Rahman M. A dynamic K-means clustering for data mining. Indonesian Journal of Electrical Engineering and Computer Science [Internet]. 2019 [citado el 13 de octubre de 2024];13(2):521. Disponible en: <http://doi.org/10.11591/ijeecs.v13.i2.pp521-526>
 24. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa L da F, et al. Clustering algorithms: A comparative approach. PLoS One [Internet]. el 1 de enero de 2019 [citado el 11 de octubre de 2024];14(1):e0210236. Disponible en: <https://doi.org/10.1371/journal.pone.0210236>
 25. Ruiz-Mori E, Ayala-Bustamante LE, Quispe-Silvestre E, Rivas-Flores RR, Burgos-Bustamante J. Cardio-Oncology in Peru: An Emerging Discipline. JACC CardioOncol [Internet]. el 1 de noviembre de 2020 [citado el 11 de octubre de 2024];2(4):671–3. Disponible en: <https://doi.org/10.1016/j.jacc.2020.10.004>
 26. Thoumi A, Bond SJ, Dotson ME, Krieger M, Garcia PJ, Ramanujam N. Policy considerations to promote equitable cervical cancer screening and treatment in Peru. Ann Glob Health [Internet]. 2021 [citado el 13 de noviembre de 2024];87(1). Disponible en: <https://doi.org/10.5334/aogh.3442>

27. Valencia-Mesías G, Rioja-Viera P, Morante-Cruz Z, Toledo-Morote Y, Neciosup-Delgado S, Gómez-Moreno H. The current situation regarding the availability and accessibility of anticancer drugs for breast cancer in the Peruvian public health systems. *Ecancermedicalscience* [Internet]. 2021 [citado el 13 de noviembre de 2024];15(1224):1–12. Disponible en: <https://doi.org/10.3332/ecancer.2021.1224>
28. Bustamante-Coronado R, Vela-Ruiz JM, Paredes-Olivares O, Carreño-Escobedo RA. Gaps in access to oncological treatment in a reference health hospital in the south of Lima in 2019. *Revista de la Facultad de Medicina Humana* [Internet]. julio de 2022 [citado el 13 de octubre de 2024];3(1):497–513. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/22506094/>

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Contribución de los autores

Conceptualización: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto

Curación de datos: Wildon Rojas Paucar, Alberto Octavio Carranza López

Análisis formal: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto

Investigación: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto

Adquisición de fondos: Wildon Rojas Paucar, Elena Miriam Chávez Garcés

Metodología: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto

Investigación: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés

Supervisión: Wildon Rojas Paucar, Alberto Octavio Carranza López

Redacción - borrador original: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto

Aprobación de la versión final: Wildon Rojas Paucar, Alberto Octavio Carranza López, Elena Miriam Chávez Garcés, Elizabeth Luisa Medina Soto