

METODOLOGÍA DE IDENTIFICACIÓN DEL GASTO PÚBLICO PARA ADULTOS MAYORES EN PERÚ: UN MODELO BASADO EN DATOS ABIERTOS Y CADENAS PRESUPUESTALES (2019-2023)

METHODOLOGY FOR IDENTIFYING PUBLIC SPENDING FOR OLDER ADULTS IN PERU: A MODEL BASED ON OPEN DATA AND BUDGET CHAINS (2019-2023)

Recibido. 5 de mayo de 2026 | Aceptado: 19 de agosto de 2026

JAVIER FRANCISCO
NAVARRETE TALAVERA¹
NELSON EDUARDO SHACK
YALTA²
ANA BERTHA COTRINA
CAMACHO³

RESUMEN

La cuantificación del gasto público peruano orientado a la población adulta mayor es imprecisa debido a las limitaciones de los clasificadores presupuestarios actuales. El objetivo de esta investigación es el diseño y especificación de un modelo procedimental para identificar y cuantificar el gasto destinado a este sector en Perú (tomando como referencia la estructura de la base de datos presupuestal del periodo 2019-2023), contrastando los clasificadores del presupuesto público con los derechos estipulados en la Ley Nro. 30490. Se establecieron cuatro niveles de identificación del gasto para adultos mayores entre los que destaca la vinculación con datasets de servicios públicos y atenciones específicas con la estructura presupuestaria vigente para estimar y calcular más recursos destinados a los adultos mayores. Asimismo, la metodología emplea un enfoque iterativo de búsqueda en cadenas presupuestales mediante términos clave y uso de consultas a un motor de inteligencia artificial. Los resultados permitieron diseñar una herramienta que identifica temáticas no visibles directamente en el presupuesto público peruano, proponiendo nuevos procedimientos para el análisis de la información. Se concluye que el modelo propuesto (que puede ser replicable a otras temáticas como por ejemplo la niñez y juventud, personas con discapacidad y otros temas) mejora la trazabilidad del gasto, permitiendo a los gestores públicos una toma de decisiones informada basada en la correspondencia real entre el presupuesto y el marco normativo de protección al adulto mayor.

Palabras clave: gasto público, clasificadores presupuestarios, adulto mayor, ley 30490, datos abiertos, big data, analítica de datos.

ABSTRACT

The quantification of Peruvian public expenditure targeted at the elderly population is imprecise due to the limitations of current budget classifications. The objective of this research is the design and specification of a procedural model to identify and quantify expenditure allocated to this sector in Peru (using the budget database structure for the 2019–2023 period as a reference), contrasting public budget classifications with the rights stipulated in Law No. 30490. Four levels of expenditure identification for older adults were established, among which the linking of public service datasets and specific care records with the current budget structure stands out, aiming to estimate and calculate additional resources allocated to older adults. Likewise, the methodology employs an iterative search approach within budget lines through key terms and the use of queries to

- 1 Magister en Gestión Pública por el Instituto de Gobierno y de Gestión Pública de la USMP
Correo: jfnavat@gmail.com
ORCID: [0009-0009-1573-6735](https://orcid.org/0009-0009-1573-6735)
- 2 Magister en Gestión y Políticas Públicas por la Universidad de Chile
Ex Contralor General de la República del Perú
Correo: shack_ne@up.edu.pe
ORCID: [0009-0002-7121-2564](https://orcid.org/0009-0002-7121-2564)
- 3 Doctorado en Gestión Universitaria
Instituto de Gobierno y de Gestión Pública de la USMP
Correo: abcotrinacamacho@gmail.com
ORCID: [0000-0003-3330-6316](https://orcid.org/0000-0003-3330-6316)

an artificial intelligence engine. The results allowed for the design of a tool that identifies topics not directly visible in the Peruvian public budget, proposing new procedures for information analysis. It is concluded that the proposed model (which can be replicated for other areas such as children and youth, persons with disabilities, and other topics) improves expenditure traceability, enabling public managers to make informed decisions based on the actual alignment between the budget and the regulatory framework for the protection of the elderly.

Keywords: public spending, budget classifiers, senior citizens, law 30490, open data, big data, data analytics.

INTRODUCCIÓN

El envejecimiento demográfico representa un desafío estructural importante para la gestión pública en el Perú, exigiendo la adaptación de los servicios estatales en pensiones, salud y cuidados y otros (MIMP, 2021). En el Perú, existe la Ley Nro. 30490 (Ministerio de la Mujer y Poblaciones Vulnerables [MIMP], 2016) también conocida como Ley de la Persona Adulta Mayor, la cual establece un marco de derechos y deberes para garantizar el bienestar de la Población Adulta Mayor (PAM). Sin embargo, la transición de la norma legal a la implementación de políticas públicas efectivas enfrenta un obstáculo crítico: la debilidad en la trazabilidad del gasto público a nivel presupuestal.

A pesar de los avances en el presupuesto por resultados (PpR), particularmente durante la primera década de este siglo, la estructura presupuestaria actual en el Perú no permite identificar de manera exhaustiva y directa la totalidad de los recursos destinados a financiar la prestación de servicios públicos a la PAM. Gran parte del gasto orientado a este sector se encuentra diluido en categorías genéricas o cadenas presupuestarias de carácter multisectorial que no cuentan con una etiqueta específica para este grupo etario. Esta “invisibilidad presupuestaria y financiera” limita la capacidad de los gestores públicos para evaluar el impacto real de las políticas de protección y dificulta la rendición de cuentas sobre el cumplimiento de la Ley 30490.

La literatura económica sobre la distribución del gasto público sugiere que, ante la falta de identificadores directos, es necesario recurrir a métodos de estimación indirecta. Al respecto, De La Fuente (2017), sostiene que es posible estimar el gasto por servicios y ubicación geográfica de forma proporcional al número de usuarios reales o potenciales, ponderando los distintos grupos poblacionales según su intensidad de uso de los servicios públicos. Bajo esta premisa, el uso de datos abiertos gubernamentales surge como una oportunidad para refinar la cuantificación del gasto mediante el cruce de información administrativa y financiera.

La presente investigación, continuando con los desarrollos de Shack (2014) y Ministerio del Ambiente [MINAM] (2015), propone un diseño de modelo procedimental basado en una herramienta de búsqueda de cadenas presupuestales que utiliza términos clave y datos asociados. La implementación de este modelo procedimental se verá potenciada mediante el uso de herramientas de inteligencia artificial y otros métodos complementarios, los que permitirán una identificación del gasto más inteligente y eficiente.

El objetivo central es el diseño y especificación del modelo procedimental (y no su aplicación cuantitativa completa al periodo 2019-2023). El modelo propuesto busca superar las limitaciones de los clasificadores actuales mediante un proceso iterativo que relacione la ejecución presupuestal con los derechos expresados en la Ley 30490. A través de la integración de *datasets* de portales de datos abiertos (como SUSALUD e INPE) y la aplicación de criterios de proporcionalidad geográfica y etaria, este estudio busca visibilizar el gasto indirecto, proporcionando indicadores de ejecución que contribuyan a una toma de decisiones informada y a la optimización de los servicios públicos para los adultos mayores en el Perú.

El presente artículo discute una taxonomía de búsqueda de información presupuestal basados en una propuesta técnica de la “Identificación del Gasto presupuestal” (Navarrete 2024). A continuación, se describen los niveles macro de la metodología que se propone.

Tabla 1
Niveles de identificación de cadenas PAM y descripción metodológica

NIVEL DE IDENTIFICACIÓN	DESCRIPCIÓN METODOLÓGICA	RECURSOS DE DATOS
Nivel 1: Directo	Identificación mediante descriptores presupuestales del Ministerio de Economía y Finanzas (MEF) (Funciones/Programas).	Estructura programática presupuestal del MEF.
Nivel 2: Contextual	Vinculación semántica unívoca o absoluta (ej. “asilos, anciano”).	Diccionario de datos presupuestales propuesto.

Nivel 3: Proporcional	Identificación y cálculo de cadenas PAM mediante inferencia estadística y ponderación de uso basado en otras fuentes de datos.	Datasets externos de datos abiertos (SUSALUD/INPE/ CONADIS/ MINSA).
Nivel 4: Residual por IA	Identificación con inteligencia artificial	Dataset de cadenas presupuestales no identificables en forma absoluta.

Nota. Elaboración propia basada en los criterios de categorización de datos.

El nivel 1 de identificación está basado en la estructura funcional programática del presupuesto público, es decir, la data que puede ser identificada directamente con los clasificadores vigentes.

El nivel 2 se refiere a la búsqueda de cadenas presupuestales que contienen términos absolutos relacionados con la PAM, es decir, términos que identifiquen a las cadenas con certeza de vinculación absoluta. Para ello se cruza la base de datos presupuestal con el diccionario de datos propuesto.

Mientras que el nivel de identificación 3, permite la desagregación de cadenas presupuestarias mediante la integración con *datasets* gubernamentales (como aquellos de SUSALUD, INPE, entre otros), aplicando factores de ponderación por UBIGEO y grupo etario de acuerdo con los datos de servicios prestados. Esto resuelve el problema de la «indivisibilidad» del gasto en servicios compartidos. Es decir, no solo se trata de una búsqueda de términos estructurados al interior de los datos de texto de las cadenas y su clasificación, sino que se plantea una fórmula de estimación presupuestaria basada en evidencia para resolver el problema de la atribución indirecta del gasto. Para esta forma de identificación se buscan las cadenas presupuestales de gasto corriente que están fuera de los Programas Presupuestales (PPTAL) del Presupuesto por Resultados (PpR) del MEF. El reto identificado fue la invisibilidad del gasto en unidades ejecutoras de atención primaria que, por naturaleza, atienden a diversos grupos etarios.

Finalmente, el grupo 4 (Residual por IA), incorpora la identificación de todas las cadenas ambiguas (con términos relativos a la PAM, pero no absolutos como “población vulnerable”) que se envían a su análisis por IA para que esta analice el contexto de la cadena y determine su relación o no con el gasto PAM.

Mediante la interoperabilidad semántica entre la base de datos (BD) del Presupuesto Público del MEF y los *datasets* de producción asistencial de SUSALUD, se normalizaron los códigos de UBIGEO del Instituto Nacional de Estadística e Informática (INEI) para permitir una distribución proporcional del gasto. Este proceso de normalización de la variable UBIGEO es importante debido a las diferencias en los *datasets* del Estado peruano, donde los datos presupuestales del MEF usan el estándar del INEI, mientras que los *datasets* de producción asistencial operan bajo la codificación del Registro Nacional de Identificación y Estado Civil (RENIEC). Aunque esto se está estandarizando, al tratarse de datos históricos es necesaria esta normalización para evitar errores de emparejamiento de datos. Esto se implementa con una matriz de correspondencia de códigos.

Como caso de estudio, se analizó una cadena presupuestal del distrito de Aramango (Región Amazonas, Provincia de Bagua). Esta cadena de gasto corriente en atención primaria tiene un monto de S/ 556,000 pero el clasificador tradicional vigente a 2025, no permite extraer el beneficio específico para la PAM. Sin embargo, al aplicar el factor de ponderación derivado de la base de datos de SUSALUD (donde se registró un 59.25% de atenciones específicas para el grupo etario ≥60 años), el modelo permitió estimar una inversión de S/ 329,430.60 destinada a este sector. Este método muestra que esta sección de la metodología propuesta permite recuperar datos de ejecución financiera que actualmente se consideran «gastos generales» e invisibles para la PAM.

El presente estudio se basa exclusivamente en la base de datos del Presupuesto Público del Perú (MEF, 2025) y fuentes de datos abiertos del gobierno del Perú por que no se tiene acceso a las bases de datos transaccionales de cada institución pública involucrada en el gasto para la PAM.

El alcance del estudio se centra en el ámbito del planteamiento y diseño del modelo procedimental que permita identificar gasto PAM desde el presupuesto público. En tal, sentido no profundiza en el análisis de los resultados cuantitativos que el método de búsqueda y cálculo plantea. Para esta finalidad, se ha analizado la estructura de la base de datos del presupuesto público en el periodo de los años 2019 a 2023. En tal sentido, solo se presentan algunas cifras a manera de ejemplo de algunos de estos años. No es el objetivo del estudio cuantificar los resultados de la aplicación del modelo.

METODOLOGÍA

La presente investigación adopta un enfoque cuantitativo y un alcance descriptivo, basado en un diseño no experimental mediante la recolección estructurada de datos (Hernández-Sampieri et al., 2022). Para ello, se desarrolló un modelo procedimental orientado a la minería y análisis de datos presupuestales del gobierno del Perú.

Inicialmente, se eliminaron, de la base datos total 2019-2023 del MEF, las cadenas presupuestales con importe de Presupuesto Institucional Modificado (PIM) e Importe de Devengado en cero que no son relevantes para el estudio y el procesamiento.

Cada nivel señalado en la Tabla 1 se estructura en fases secuenciales que se describen a continuación.

Nivel 1: Identificación mediante descriptores asociados (clasificadores).

- Identificación de programas presupuestales exclusivos para el gasto PAM en el presupuesto público.
- Identificación de funciones, divisiones, grupos funcionales y componentes en campos categóricos relativos al gasto en PAM.
- Identificación de Pliegos Presupuestales que ejecutan gasto PAM exclusivo.
- Separación de *dataset* de registros identificados y marca de cadenas usadas.
- Asignación del identificador correspondiente a la temática hallada de acuerdo con el clasificador de gastos PAM propuesto (CLAGPAM) de la tabla 2.

Nivel 2: Identificación directa en base a términos de diccionario temático.

- Definición de términos absolutos y relativos del diccionario de búsqueda.
- Diseño del algoritmo de identificación de cadenas de gasto PAM.
- Separación de *dataset* de registros identificados y marca de cadenas usadas.
- Asignación del identificador correspondiente a la temática hallada de acuerdo con el clasificador (CLAGPAM) de la tabla 2.

Nivel 3: Identificación y cálculo de importes de cadenas PAM mediante inferencia estadística y ponderación de uso basado en otras fuentes de datos.

- Identificación de *datasets* de gastos corrientes en servicios públicos en los portales de datos abiertos del Gobierno peruano.
- Descarga, conversión y uniformización de datos.
- Diseño de algoritmo de relación de cadenas de gasto corriente relativo a la PAM con los registros de servicios públicos. Cálculo y prorrateo en base a la cantidad de atenciones por grupo etario y UBIGEO distrital (PAM y no PAM).
- Separación de *dataset* de registros identificados y marca de cadenas usadas.
- Asignación del identificador correspondiente a la temática hallada de acuerdo con el clasificador de la Tabla 2.

Nivel 4: Identificación con inteligencia artificial de cadenas ambiguas para gasto PAM

- Diseño del algoritmo de identificación de cadenas de gasto PAM no absolutas en base al diccionario de términos.
- Diseño del algoritmo de agrupación y envío de cadenas presupuestales a un motor de IA local (para procesar grandes volúmenes de datos sin restricciones) para determinar si corresponden a gasto PAM por el contexto de la cadena y sus campos de texto descriptivos.
- Separación de *dataset* de registros identificados y marca de cadenas usadas.
- Asignación del identificador correspondiente a la temática hallada de acuerdo con el clasificador CLAGPAM.

A continuación, se describe cada nivel en detalle.

Nivel 1: Identificación mediante descriptores asociados (clasificadores)

Se aplicó la técnica de focalización programática directa (MEF, 2021). De esta forma, se examinaron las estructuras del presupuesto por resultados (PpR) para seleccionar únicamente los programas presupuestales cuyo diseño funcional, lógica de intervención y matriz de marco lógico estén dirigidos de manera directa y exclusiva a la atención a la PAM.

Luego, para obtener aquellas cadenas presupuestales no agrupadas en PPTALes exclusivos para la PAM, se recurrió al análisis mediante la Clasificación Funcional del Gasto (BID, 2020). La metodología consistió en realizar un filtrado jerárquico en varios niveles a través de los clasificadores categóricos como la función, división funcional, grupo funcional y la actividad/proyecto/obra), seleccionando aquellas categorías que abarcan servicios o acciones orientadas a la protección social, salud gerontológica e inclusión de la PAM.

Finalmente, dado que las pensiones de jubilación están dirigidas a adultos mayores cesantes o jubilados, se incluye en la identificación de cadenas de gasto, aquellas que corresponden al presupuesto de la Oficina de Normalización Previsional (ONP, 2024) que se encarga de este tema. Una vez concluida la identificación de cada cadena presupuestal con los criterios señalados se procede a la separación de *dataset* de registros identificados y marca de cadenas usadas para optimizar el procesamiento posterior.

Nivel 2: Identificación directa

Para aplicar una estrategia de filtrado categórico estricto (claves absolutas) y de coincidencia de patrones contextuales (claves relativas) se requiere la definición de un diccionario de términos.

La literatura científica confirma que la minería de datos sobre grandes *datasets* institucionales requiere un enfoque estructurado de selección de palabras clave basado en diccionarios de términos (Macanovic, 2022). Según Noh et al. (2015), la creación de diccionarios específicos por dominio permite un procesamiento semántico más claro a través de expresiones booleanas y clasificadores directos.

Se construyó un diccionario con términos con base en la Ley Nro. 30490. Estos términos se agruparon en dos dimensiones:

- Directos o absolutos: conceptos que aluden exclusivamente a la PAM (ej. «Centros de Atención Residencial para Adultos Mayores», «Asilos», «Vejez»).
- Indirectos/Temáticos: servicios generales que incluyen a la PAM (ej. «Atención ambulatoria», «Asistencia alimentaria»).

En la minería de datos presupuestales y otras materias, la estructuración de variables en dimensiones absolutas (directas) y relativas (indirectas) encuentra su paralelo técnico en los modelos multietapa impulsados por léxicos. Al respecto, Macanovic (2022) señala que las metodologías avanzadas de minería de texto y el análisis computacional ofrecen un panorama transformador para el análisis de datos masivos en las ciencias sociales.

La búsqueda de términos clave absolutos (lexemas unívocos) del diccionario definido sobre campos de texto en la base de datos del presupuesto público se fundamenta en los modelos de extracción de información basados en diccionarios deterministas. En el marco del análisis computacional de texto, Grimmer y Stewart (2013) demuestran que los enfoques basados en diccionarios constituyen herramientas fundamentales para la medición automatizada de variables conceptuales complejas. Trasladado a conjuntos de datos masivos (como el presupuesto público), este proceso de coincidencia de palabras clave, se consolida como un enfoque de clasificación altamente reproducible, transparente y metodológicamente robusto (Grimmer y Stewart, 2013). Al operar bajo lógicas deterministas, este tipo de diseño algorítmico se orienta prioritariamente a mitigar el riesgo, buscando reducir a cero los falsos positivos en la asignación de etiquetas de categoría a los datos públicos.

El algoritmo se basa en la construcción de expresiones regulares (cadenas de texto que unen a los términos del diccionario) con el operador lógico “|” (OR) para maximizar la performance de las búsquedas de varios términos en una sola iteración. Luego, se iteran las búsquedas en cada registro presupuestal en los campos categóricos y de texto siguientes: programa presupuestal, división, grupo funcional, finalidad del gasto (meta del gasto), específica de gasto y detalle de específica de gasto. Se diseñó un proceso iterativo de filtrado que opera sobre los campos de texto del *dataset* del Presupuesto Público. Técnicamente, este procedimiento se sustenta en los modelos de extracción de información basada en reglas y coincidencia léxica determinista. El algoritmo ejecuta búsquedas de coincidencia de lexemas (raíces de palabras de mínimo 4 letras) exacta a partir del diccionario de términos clasificados como absolutos previamente estructurado en el diccionario, etiquetando las cadenas presupuestales que se van identificando. La aplicación de búsquedas basadas en diccionarios de palabras clave sobre datos masivos funciona de manera óptima en el procesamiento de textos, permitiendo reducir el sesgo humano al estandarizar los criterios de búsqueda bajo reglas algorítmicas neutrales y explícitas (Reveilha & Morselli, 2022).

Asimismo, la literatura en minería de datos demuestra que la búsqueda basada en diccionarios de términos clave es altamente efectiva para trabajar con bases de datos masivas. Al mapear y asociar este diccionario estructurado a campos específicos de un gran volumen de datos, el sistema puede descartar eficazmente los registros irrelevantes (ruido informativo). Esto optimiza el procesamiento algorítmico y facilita la segmentación de los datos en categorías temáticas altamente granulares (Li y Li, 2024).

Las cadenas que se van identificando se marcan como usadas para optimizar las futuras búsquedas. Al finalizar el proceso se separa el *dataset* de registros identificados y se suma al *dataset* del nivel 1 de identificación.

Asimismo, se efectuó un análisis comparativo entre los ejes temáticos de la Ley 30490 (Ley PAM) para revisar el nivel de cobertura de los clasificadores vigentes frente a la identificación por clasificadores presentes en el presupuesto. Estos ejes son:

- Salud integral
- Seguridad social / Pensiones
- Trabajo
- Educación / Cultura
- Recreación y deporte
- Participación y organización
- Accesibilidad y movilidad
- Servicios especializados (centros de atención con servicios múltiples).
- Bienestar general y atención integral.

Nivel 3: Integración con *datasets* de otras instituciones y prorrateo del gasto

Esta fase constituye el componente metodológico central del estudio para abordar las cadenas de gasto no exclusivas (es decir, aquellas partidas presupuestales compartidas con grupos etarios donde la PAM es beneficiaria o usuaria del servicio, pero el gasto no está asignado en un 100% a este grupo de edad). Para la estimación del gasto, se adoptó la metodología de prorrateo por coeficientes de distribución y proporciones de demanda asistencial, adaptada de las pautas aplicadas por el Fondo de las Naciones Unidas para la Infancia (UNICEF, 2020) para el análisis de gasto público en niños, niñas y adolescentes. El procesamiento analítico e interoperabilidad de datos abiertos se estructuró de la siguiente forma:

1. Extracción: descarga de datos detallados del presupuesto público y *datasets* de producción asistencial de la Superintendencia Nacional de Salud (SUSALUD), Instituto Nacional Penitenciario (INPE), MINSA (Ministerio de Salud) y el Consejo Nacional para la Integración de la Persona con Discapacidad (CONADIS).
2. Normalización: homogenización de la variable UBIGEO (alineando con el estándar INEI usado por el Ministerio de Economía y Finanzas).
3. Agrupación y totalización del gasto corriente de los servicios para la PAM y no PAM.
4. Cálculo del Factor de Atribución (FA): se calcula la proporción de atención específica para el grupo etario de mayores de 60 años en cada unidad territorial:

$$F_a = \frac{\text{Atenciones PAM}}{\text{Total Atenciones (PAM + No PAM)}} * 100$$

5. Estimación del gasto: aplicación del FA al monto total de la cadena presupuestal para determinar la inversión en la PAM.

La programación del proceso se ha efectuado en el lenguaje de programación Python, empleando librerías de inteligencia artificial como la biblioteca Scikit-learn (Pedregosa et al., 2011) para el preprocesamiento de Machine Learning y otras librerías que incluyen Pandas y NumPy para las manipulaciones numéricas, y SciPy para las transformaciones estadísticas y escalar datos a rangos específicos o estandarizarlos. Dado que el presente estudio está orientado a lectores afines a la gestión pública, el presente documento no incluye el código fuente y los *scripts* de programación y se centra en la descripción lógica, funcional y conceptual de la arquitectura metodológica.

Para todas las cadenas presupuestales (fuera del PpR) que incluyen dentro del contenido de sus campos descriptivos a algún término que se refiera a servicios o apoyo a poblaciones vulnerables que involucren a la PAM de manera no exclusiva, se plantea un método específico para distribuir el gasto que contiene la cadena en base al criterios de prorrateo de porcentaje atenciones en salud por grupo de edad en el área geográfica o ámbito de acción del pliego presupuestario o unidad ejecutora responsable del gasto corriente.

Nivel 4: Residual por IA

Debido al volumen masivo del *dataset* presupuestal (incluso considerando el *dataset* reducido con todos los registros que fueron marcados como usados en los niveles anteriores) se estima que quedan por analizar un grupo muy numeroso de cadenas (más de 500 mil en el periodo de estudio) que corresponden a asistencia y protección social en general. Enviar dichas cadenas a una IA online presenta serias limitaciones operativas tales como restricciones en las cuotas diarias de consultas, costos altos por uso de la IA en forma masiva y tiempos

amplios de transferencia de datos (Chen et al., 2023). Por esta razón se plantea el uso de un motor de IA que se instala localmente (Llama 3) y que se integra con Python para permitir el procesamiento de los registros de manera más rápida y sin la restricción de la cantidad de *prompts* (consultas a la IA) al día que permite una IA *online* a un usuario. De tal forma, se busca que las cadenas sean categorizarlas automáticamente.

“El empleo de arquitectura de lenguaje abierta alojada localmente responde a la necesidad de realizar consultas ilimitadas y procesamiento concurrente de bases de datos masivas sin depender de servidores externos ni tarifas por uso de API⁴ (Meta AI, 2024).”

Tabla 2
Clasificador propuesto para cuantificar el gasto en PAM (CLAGPAM)

TEMA PRINCIPAL	SUBTEMA / PROGRAMA	PRODUCTO / PROYECTO	ACTIVIDAD / OBRA (PRESUPUESTO POR RESULTADOS)	DERECHO EN LEY PAM
1. Salud integral (gasto corriente)	1.1 Salud general 1.2 Salud mental	–	–	Una vida saludable. Atención integral en salud y participar del proceso de atención de su salud.
2. Seguridad Social y pensiones	2.1 Pensiones ONP	2.1.1 Pensiones 2.1.2 Indemnizaciones	–	Una vida digna.
	2.2 PPTAL Pensión 65	2.2.1 Acciones comunes	2.2.1.1 Gestión del Programa	Una vida digna.
		2.2.2 Adulto mayor con subvención monetaria	2.2.2.1 Entrega de la subvención monetaria a los beneficiarios 2.2.2.2 Afiliación y verificación de requisitos 2.2.2.3 Visitas de identificación y soporte a usuarios con vulnerabilidad	Una vida digna.
	2.3 Otras obligaciones previsionales	–	–	Una vida digna.
3. Trabajo (gasto corriente)	–	–	–	La igualdad de oportunidades. Realizar labores o tareas acordes a su capacidad física o intelectual.
4. Educación y cultura (gasto corriente)	–	–	–	La igualdad de oportunidades. Acceder a programas de educación y capacitación.
5. Recreación y Deporte	5.1 Espacios públicos amigables	5.1.1 Áreas verdes 5.1.2 Áreas deportivas	–	Una vida digna y plena.
	5.2 Clases diversas	–	–	Una vida digna y plena.

4 API: Application Programming Interface (Interfaz de Programación de Aplicaciones).

METODOLOGÍA DE IDENTIFICACIÓN DEL GASTO PÚBLICO PARA ADULTOS MAYORES EN PERÚ: UN MODELO BASADO EN DATOS ABIERTOS Y CADENAS PRESUPUESTALES (2019-2023)

6. Participación y Organización (gasto corriente)	–	–	–	Participar activamente en las esferas social, laboral, económica, cultural y política del país.
7. Accesibilidad, movilidad y transitabilidad	–	–	–	Atención preferente en establecimientos públicos. Una vida independiente, autónoma.
8. Acceso a servicios especializados	8.1 Servicios especializados para la PAM	8.1.1 Acciones comunes	8.1.1.1 Gestión del Programa 8.1.1.2 Seguimiento y evaluación del programa	Recibir atención integral e integrada, cuidado y protección familiar y social.
		8.1.2 Familiares y cuidadores con capacidades fortalecidas en el buen trato al AM	8.1.2.1 Fortalecimiento de las capacidades a los operadores del programa 8.1.3.1 Desarrollo de competencias en familiares para la atención de PAM	Recibir atención integral e integrada, cuidado y protección familiar y social.
		8.1.3 PAM atendidas involucrando al entorno familiar y social	8.1.3.2 Identificación, selección y derivación de PAM en situación de riesgo 8.1.3.3 PAM en situación de riesgo 8.1.3.4 PAM reciben servicios para prevenir condiciones de riesgo 8.1.3.5 Acreditación y supervisión de centros de atención para PAM 8.1.3.6 Medidas de protección temporal para PAM en situación de riesgo	Recibir atención integral e integrada, cuidado y protección familiar y social.
		8.1.4 Creación, mantenimiento o ampliación de centros de atención para AM	8.1.4.1 PAM en situación de riesgo atendidas en CAN 8.1.4.2 PAM en situación de riesgo atendidas en CAR 8.1.4.3 PAM en situación de riesgo atendidas en CAD	Recibir atención integral e integrada, cuidado y protección familiar y social.
9. Reclusión (gasto corriente)	–	–	–	Acceder a condiciones apropiadas de reclusión. Acceso a la justicia.
10. Bienestar en general	10.1 Vida digna 10.2 Adicciones 10.3 Otros	–	–	Una vida digna, plena. No discriminación por edad.

Nota. Adaptación basada en el clasificador propuesto en Navarrete (2024). Las celdas marcadas con una línea (–) indican que el nivel jerárquico presupuestal no cuenta con una apertura específica en esa categoría.

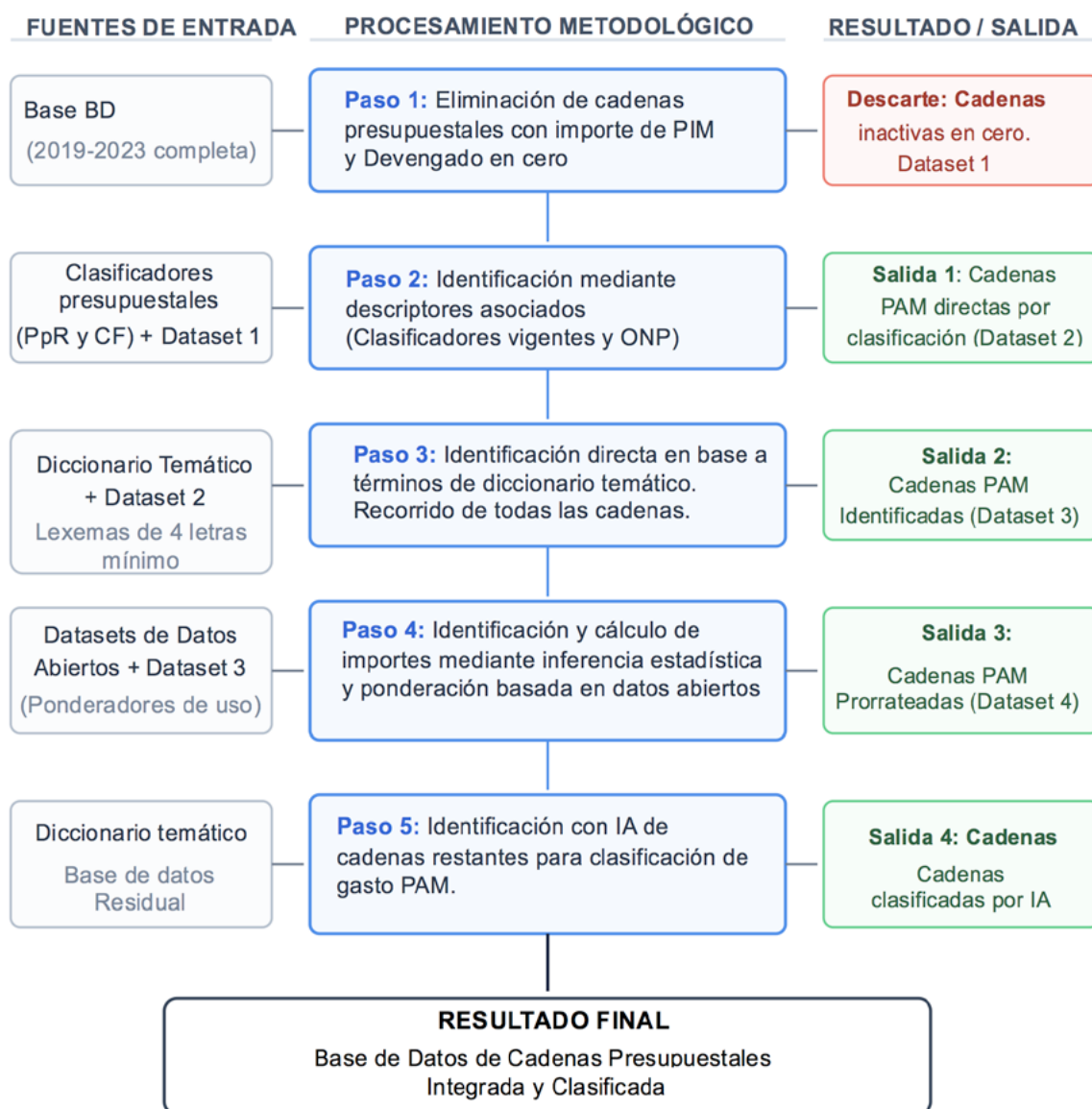
RESULTADOS

Dado que el presente artículo propone un modelo de análisis y cálculo en lugar de métricas de ejecución o resultados numéricos, en esta sección se describe la especificación formal del modelo creado y los hallazgos producto del análisis de datos categóricos.

El modelo general macro

Figura 1

Diagrama de flujo metodológico para el procesamiento y clasificación de cadenas presupuestales PAM



Nota. Elaboración propia.

Todos los pasos de la Figura 1 incorporan también el CLAGPAM (para que el proceso asigne una entrada a cada cadena para el clasificador de gastos).

Diseño de entidades de datos y especificación detallada

• Clasificadores presupuestales (MEF)

Tabla 3

Clasificación Presupuestal del Presupuesto por Resultados y Funcional-Programático

CLASIFICADOR PRAGMÁTICO PRESUPUESTO POR RESULTADOS (PPR)			CLASIFICADOR FUNCIONAL				
Categoría Presupuestal	Producto/ Proyecto	Actividad/Obra/ Acción de inversión	Función	División Funcional	Grupo Funcional	Componente	Finalidad
Programa Presupuestal (PPTAL)	Producto	Actividad	Funciones como Ambiente, Salud, Educación	Divisiones de la Función como: Saneamiento, Gestión integral de la calidad ambiental, Asistencia social, Desarrollo urbano, etc.	Agrupa las divisiones funcionales. Por ejemplo: Promoción laboral, Saneamiento rural, Protección de poblaciones en riesgo, etc.	División que permite identificar acciones presupuestales específicas. Es el contenedor de las metas presupuestales. Algunos componentes son: Apoyo comunal, Desarrollo económico y social, Fortalecimiento de capacidades, etc.	Detalle en texto de la finalidad del gasto de la cadena presupuestal. Algunas descripciones son genéricas y otras específicas.
Acciones centrales	Proyecto	Obra					
Asignaciones Presupuestarias que no resultan en Productos (APNOP)		Acción de inversión					

Nota. Ministerio de Economía y Finanzas 2025 (MEF). Relación y Estructura Programática de los Programas Presupuestales.

Se identificaron dos programas presupuestales exclusivos para la PAM. La ejecución total devengada de estos programas, Pensión 65 y el programa de Servicios Especializados para Adultos Mayores entre los años 2019 y 2023 se detalla el resultado a continuación porque se obtiene directamente del presupuesto sin necesidad de ejecutar el modelo. Los datos reflejan el gasto financiero ejecutado en millones de soles (S/) según los registros de la Consulta Amigable del Ministerio de Economía y Finanzas (MEF) y balances institucionales oficiales.

Tabla 4

Ejecución presupuestal de los programas Pensión 65 y Servicios Especializados para adultos mayores (2019-2023)

PROGRAMA PRESUPUESTAL	2019	2020	2021	2022	2023
0097: Pensión 65	859.5	5,579.0	884.2	1,061.2	1,356.9
0142: Servicios Especializados	20.0	24.0	36.9	49.8	86.0
Total	879.5	5,603.0	921.1	1,111.0	1,442.9

Nota. Valores expresados en millones de soles (S/). Los datos reflejan el presupuesto ejecutado. Extraído de la Consulta Amigable del (MEF) del Perú.

El PPTAL 0142 abarca 59 productos/proyectos relacionados con infraestructura y servicios de cuidados en Centros de atención al adulto mayor: CIAM: Centro Integral de Atención al Adulto Mayor, CAN: Centro del Adulto Mayor, CAR: Centro de Atención Residencial y CAD: Centro de Atención de Día (servicios que brindan cuidado durante el día sin que la persona adulta mayor resida en el lugar).

A pesar de la trazabilidad directa de estos programas, se observa que su alcance es limitado, centrándose exclusivamente en protección social y asistencia, dejando fuera componentes críticos como salud especializada, trabajo o accesibilidad.

Los resultados agregados muestran comportamientos diferenciados en términos de eficacia en la ejecución presupuestaria:

Pensión 65 (PPTAL 097): Presenta el desempeño más alto con una ejecución promedio del 99.61 % (Nivel: Muy Bueno, según escala MEF), consolidando un gasto total de S/ 4,581 millones en el periodo de estudio.

Servicios Especializados (PPTAL 142): Muestra una tendencia creciente, pasando de S/ 30 millones en 2019 a S/ 74 millones en 2023. Su nivel de ejecución promedio (86.79%) se sitúa en una valoración de “Regular” a “Bueno” según la escala de valoración del MEF (MEF, 2022).

Componente “Apoyo al Anciano”: Fuera de los programas estratégicos, este rubro del Clasificador Funcional ejecutó S/ 10.9 millones, con una eficacia promedio del 87.47%.

Gestión Previsional (ONP): Representa el volumen monetario más amplio, con una ejecución quinquenal de S/ 37,876 millones.

• Diccionario de términos de búsqueda

Las cadenas asociadas al gasto PAM identificadas en el paso 2 de la Figura 1 (identificación por clasificadores vigentes y el gasto previsional de la ONP) fueron contrastadas con los ejes de la Ley 30490, encontrándose una coincidencia de solo 22.27%. Ante esta limitación de los clasificadores presupuestales, la presente investigación propone un modelo basado en una minería de datos presupuestarios desde un diccionario de términos clave y los niveles antes señalados.

A diferencia de una búsqueda simple, la búsqueda de términos por diccionario utiliza técnicas de truncamiento para maximizar la recuperación de información.

Optimización de búsqueda: El uso de raíces léxicas (ej. *JUBILA*, *GERIATRI*) permite capturar variaciones morfológicas (género, número y derivaciones) que los filtros estáticos del MEF omiten.

Jerarquización de llaves: El algoritmo opera bajo un principio de especificidad descendente. Se priorizan términos de alta precisión temática (ej. *ASILO*, *CASA DE REPOSO*) antes que términos más generales (ej. *SALUD*, *BIENESTAR*), reduciendo así el ruido en la clasificación y el riesgo de falsos positivos.

El modelo no solo cuantifica dinero, sino que traduce el gasto a derechos. Cada llave del diccionario está vinculada a un Identificador de Tema (ID) dentro del CLAGPAM propuesto en esta investigación.

Tabla 5

Ejemplos de llaves del Diccionario de Términos de Búsqueda con asociación al CLAGPAM.

TIPO DE LLAVE	EJEMPLOS DE TÉRMINOS	ASOCIACIÓN CON CLAGPAM
Específica clínica	GERIATRI*, CATARATA, DEMENCIA	Salud Integral
Institucional/Cuidado	ASILO, CASA DE REPOSO, CIAM	Servicios Especializados
Socioeconómica	JUBILA*, PENSION, ABANDONO	Seguridad Social / Protección
Transversal	ACCESIBILIDAD, MOVILIDAD, EMPLEO	Derechos de Autonomía

Nota. Elaboración propia.

Los términos o llaves de búsqueda se han categorizado de la siguiente forma:

Categoría de términos absolutos (Identificadores directos PAM)

- ADULTO MAYOR / ADULTOS MAYORES / ADULTA MAYOR
- CASA DE REPOSO / ASILO
- TERCERA EDAD / VEJEZ / ENVEJECIMIENTO
- ANCIAN (anciano, anciana, ancianato)
- GERIATR (geriatría, geriátrico)
- GERONTOL (gerontología, gerontólogo)
- LONGEV
- Centro de Atención Residencial
- PENSIÓN / CESANTE / JUBILA (jubilado, jubilación)
- Centro de Atención de Noche
- Centro de Atención Diurna / Centro de Atención de Día
- CIAM (Centro Integral de Atención al Adulto Mayor)
- CARPAM (Centro de Atención Residencial para Personas Adultas Mayores)

Categoría 2: Identificadores indirectos PAM

Deben buscarse cerca de palabras de la categoría de absolutos para evitar falsos positivos.

- RECLUSION / PENAL / PENITENCIARI / CARCEL
- DEMENCIA / ALZHEIMER
- DEGENERATIV

- CÁNCER / CEGUERA / CATARATA
- VACUNACIÓN / NEUMOCOCO.
- INTERVENCIÓN QUIRÚRGICA / ATENCIÓN MED / OPERACIÓN
- HOSPITAL/ POSTA / MEDIC/ CENTRO DE SALUD
- OSTEOPOROSIS /PARKINSON
- RAMPA / PASAMANOS / BARANDA / SILLA DE RUEDAS
- ABANDONO / VULNERA / BENEFICENCIA
- SUBVENCION MONETARIA
- TRABAJO / EMPLEO
- CUIDADOR
- BIENESTAR / RECREACIÓN
- CULTURA / ALFABETIZA / DEPORT
- ASESORAMIENTO JURÍDICO / ASESORAMIENTO LEGAL
- IGUALDAD / ACCESIBILIDAD / TRANSITABILIDAD

Los términos y lexemas del diccionario son agrupados en expresiones regulares (cadenas de texto separadas por el operador | que significa OR), para que las búsquedas se realicen en forma masiva comparando los datos de los campos presupuestales como una expresión regular compuesta para minimizar la búsqueda individual o iterativa por cada término. Ejemplo: “asilo | adulto mayor | vejez | tercera edad”.

Tabla 6

Estructura secuencial del algoritmo de depuración y compilación de patrones del Diccionario de Búsqueda

ETAPA DEL PROCESO	OPERACIÓN / FUNCIÓN DE CÓDIGO	ENTRADA Y VALIDACIÓN LÓGICA	SALIDA Y RESULTADO ESPERADO
1. Carga	Extracción SQL (read_sql)	Consulta a Diccionario PAM	DataFrames de términos léidos.
2. Depuración I	Eliminación de duplicados (fromkeys)	Mantenimiento del orden secuencial	Listas sin redundancias.
3. Bifurcación I	Control de longitud de lexemas (len <= 4)	Raíces cortas vs. Raíces largas	Evaluación del riesgo de falsos positivos.
4. Depuración II	Filtro de seguridad (len > 1)	Exclusión de caracteres únicos o nulos	Listas definitivas optimizadas (absolutos y relativos).
5. Compilación	Agrupamiento óptimo	Unión mediante operador OR ()	Objetos Regex en caché.

Nota. Elaboración propia. El diagrama ilustra las dos fases consecutivas de normalización léxica y construcción dinámica de expresiones regulares aplicadas al diccionario de la PAM.

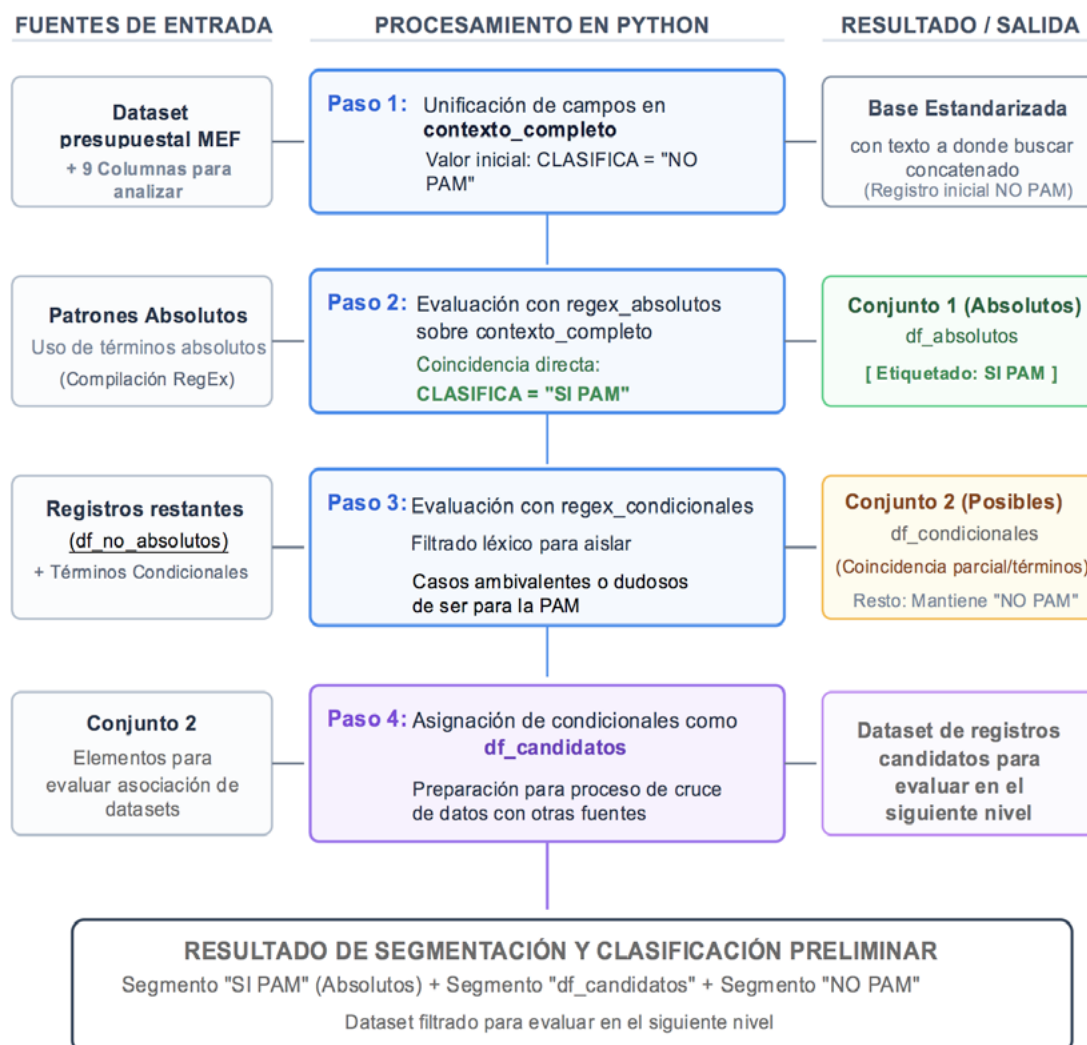
Para buscar en los campos del presupuesto se ha seleccionado los siguientes campos (columnas) de texto descriptivo de las cadenas presupuestales:

1. Programa_Ppto_Nombre (nombre del PPTAL).
2. Producto_Proyecto_Nombre (nombre de Producto/Proyecto).
3. Actividad_Accion_Obra_Nombre (nombre de la Acción/Obra).
4. Funcion_Nombre (nombre de la Función).
5. Division_Funcional_Nombre (nombre de la División Funcional).
6. Grupo_Funcional_Nombre (nombre del Grupo Funcional).
7. Meta_Nombre (descripción de la Finalidad del Gasto).
8. Categoria_Gasto_Nombre (nombre de la categoría de gasto).
9. Sub_Generica_Det_Nombre (descripción detallada de la genérica de gasto).

Con los patrones de búsqueda armados y con los campos del presupuesto sobre los cuales aplicar la búsqueda se procede con la descripción del flujo del procesamiento basado en el diccionario de términos.

Figura 2

Diagrama de flujo para el procesamiento y clasificación de cadenas presupuestales PAM en base al Diccionario de Términos



Nota. Elaboración propia.

El diagrama refleja la lógica del código de *script* en Python. Unifica campos cualitativos en la columna `contexto_completo` y efectúa una inicialización general en el valor "NO PAM" para las cadenas no usadas aún. Luego, se ejecuta un proceso iterativo para evaluar los patrones absolutos para clasificar las cadenas presupuestales y colocarlas en el dataset "df_absolutos" marcadas como "SI PAM". Luego, se evalúan los registros presupuestales en base a los patrones condicionales en el conjunto de datos restante (`df_no_absolutos`) para identificar candidatos ambivalentes. Y finalmente, se estructura y segmenta el conjunto "df_candidatos" listo para ser enviado a su evaluación por comparación con datasets de otras instituciones públicas que brindan servicios para la PAM.

Identificación de cadenas PAM de servicios públicos relativos a *datasets* de otras instituciones.

Para la identificación de gasto PAM adicional se utilizaron bases de datos de consulta pública las entidades: SUSALUD, INPE, MINSA (Salud mental) y CONADIS que pueden asociarse geográficamente y por edades a las cadenas presupuestales que contienen datos de UBIGEO. Para obtener los datos necesarios de SUSALUD, se deben filtrar los registros de gasto corriente en salud en general por un distrito (UBIGEO) agregando los datos PAM y No PAM. De esta forma, se podrá efectuar una distribución porcentual en base al porcentaje de PAM atendida en el distrito según los datos de atenciones SUSALUD por grupo de edad y UBIGEO.

Para ello se descargan los datasets en Excel o formato CSV con el detalle de las atenciones de la población por UBIGEO (campo `ATE_UBIGEO`) y edad o rango de edad (campo `ATE_EDAD`) como se muestra a continuación:

Tabla 7

Muestra de datos del dataset descargado de atenciones ambulatorias de SUSALUD 2019-2023

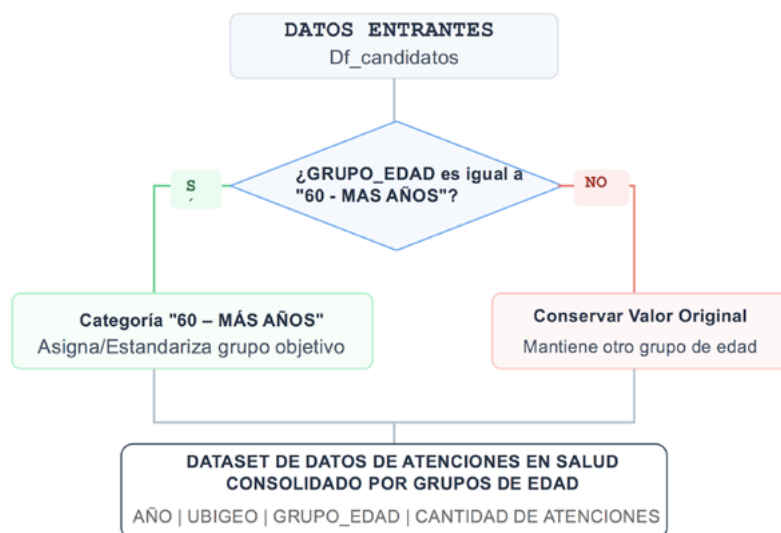
MES DE ATENCIÓN	ID DE ATENCIÓN	EDAD DEL PACIENTE	SEXO (0=MASCULINO, 1=FEMENINO)	FECHA DE ALTA HOSPITALARIA (AAAA-MM-DD)	COD. UBIGEO
5	907120635	37	1	2023-05-20	120421
6	914599050	37	1	2023-05-20	120421
6	914598915	37	1		120421
8	930120344	11	0	2023-05-20	120421
11	948230157	14	1	2023-05-20	120421
12	957107442	68	1	2023-02-06	120421

Nota. Adaptado de consulta web: Producción asistencial en consulta ambulatoria, por Superintendencia Nacional de Salud (SUSALUD), 2019-2023.

La Tabla 7 presenta una muestra de los campos de los registros de atención ambulatoria de pacientes a nivel nacional. Las variables fundamentales corresponden a la edad del paciente (ATE_EDAD), utilizada para la categorización dicotómica entre PAM y no PAM, y el código de ubicación geográfica (ATE_IDUBIGEO) para establecer relaciones con la base de datos presupuestal. Mediante la consolidación de los abiertos de SUSALUD para el periodo de estudio (2019-2023), se estructuró un conjunto de datos agregado por los grupos etarios de corte (60 años a más y otros) con el fin de estimar las proporciones del gasto corriente en salud. Posteriormente, esta base de datos agregada se vinculó con la con la base de datos del Presupuesto Público utilizando la variable UBIGEO y año de ejecución como llaves de relación. Este proceso permite asociar los registros asistenciales con las cadenas presupuestales correspondientes, incorporando datos de salida presentes en dichas cadenas como el monto presupuestado, el monto ejecutado, el sector, el pliego, la unidad ejecutora, la función, el grupo funcional y la categoría de gasto, entre otros.

Figura 3

Diagrama de flujo para la clasificación de datos cantidades de atenciones en servicios de salud por grupo de edad y UBIGEO



Nota. Elaboración propia.

La data de SUSALUD agregada contiene registros por grupos de edad, por lo que esta fuente de datos debe ser filtrada para la población mayor o igual a 60 años de acuerdo con la definición PAM en la Ley. De esta forma deben agruparse las cantidades de atenciones por UBIGEO y grupo de edad ≥ 60 y < 60 . Así quedan clasificados con el campo GRUPO_EDAD = PAM toda la población objetivo y como NOPAM el resto de la población atendida en un distrito. Es decir, la tabla resultante procesada contendrá 2 registros por cada distrito del Perú. Así, por ejemplo, si la cantidad de PAM del distrito de ARAMANGO, provincia de BAGUA, departamento de AMAZONAS con

atenciones en salud pública anual es de 14,904 personas y el resto de la población representa 10,247 atenciones anuales, se puede obtener una PAM atendida estimada del 59.25%

Con este dato para cada entrada relacionada a una cadena se podrá calcular un gasto aproximado para la PAM. Por ejemplo, si en el distrito ARAMANGO /BAGUA/ AMAZONAS se atiende a personas de todas las edades con un presupuesto para gasto corriente de S/ 556,000 entonces se podrá aproximar que el monto presupuestado para la PAM de ese distrito es de S/ 329,430.60.

De manera similar al tema de salud, el gasto corriente penitenciario puede ser estimado con la ayuda de los datos abiertos publicados por el INPE (2024) en su portal web que señala una población penal de PAM de 6,154 personas 2024.

Tabla 8

Muestra de registro de población penitenciaria (INPE – Tumbes)

ID	DELITO ESPECÍFICO	SITUACIÓN JURÍDICA	EDAD	SEXO	PAÍS	EST. CIVIL	DPTO.
1	ROBO AGRAV.	SENTENCIADO	30	MASCULINO	PERÚ	SOLTERO	Tumbes
2	HOMICIDIO C.	SENTENCIADO	28	MASCULINO	PERÚ	SOLTERO	Tumbes
3	TRÁFICO ILÍC.	SENTENCIADO	23	MASCULINO	COLOMBIA	CONVIVIENTE	Tumbes
4	VIOLACIÓN SE	PROCESADO	38	MASCULINO	PERÚ	CONVIVIENTE	Tumbes
5	ROBO AGRAV.	PROCESADO	19	MASCULINO	PERÚ	SOLTERO	Tumbes
6	VIOLACIÓN SE	PROCESADO	43	MASCULINO	PERÚ	CASADO	Tumbes
7	VIOLACIÓN SE	PROCESADO	37	MASCULINO	PERÚ	CONVIVIENTE	Tumbes
8	HURTO AGRAV.	SENTENCIADO	39	MASCULINO	PERÚ	SOLTERO	Tumbes

Nota. Datos extraídos del registro de población penitenciaria (portal-inpe.opendata.arcgis.com)

Entonces, es posible agrupar la población penitenciaria por edades para separar la PAM y la cantidad No PAM por región o departamento (no se cuenta con el UBIGEO completo) y de esta forma estimar la proporción del gasto corriente penitenciario para aproximar el gasto PAM en función al porcentaje de PAM e por departamento de la cadena presupuestaria.

Respecto a los datos de atenciones en salud mental con el detalle de atenciones por UBIGEO y grupos de edad se han encontrado data en los registros de SUSALUD. En este caso se puede usar la proporción de atenciones para PAM del referido año para calcular un factor de proporción. Para ello se ha filtrado la base de datos de los servicios de salud prestados para adultos mayores por los siguientes códigos de diagnóstico (CIE correspondiente al campo VPRIDIAGCIE_ATE del dataset):

- F00.1 Demencia en la enfermedad de Alzheimer, de comienzo tardío.
- F00.2 Demencia en la enfermedad de Alzheimer, atípica o mixta.
- F00.9 Demencia en la enfermedad de Alzheimer, no especificada.
- G30.1 Enfermedad de Alzheimer de comienzo tardío.
- G30.8 Otros tipos de enfermedad de Alzheimer
- G30.9 Enfermedad de Alzheimer, no especificada
- F69.X Trastorno de la personalidad y del comportamiento en adultos.
- F41.2 Trastorno mixto de ansiedad y depresión.
- F06.4 Trastorno de ansiedad.
- F40.9 Trastorno fóbico de ansiedad, no especificado
- F41.1 Trastorno de ansiedad generalizada.
- F41.2 Trastorno mixto de ansiedad y depresión
- F41.3 Otros trastornos de ansiedad mixtos
- F41.8 Otros trastornos de ansiedad especificados
- F41.9 Trastorno de ansiedad, no especificado

La información disponible sobre atenciones en salud mental, que presenta una alta incidencia en personas adultas mayores, contiene la ubicación de cada atención por UBIGEO y el campo llamado “VFECNACPAC_ATE” que

corresponde a la fecha de nacimiento del paciente y que permite calcular su edad para determinar el rango de edad (PAM o No PAM) de las personas atendidas.

De esta forma se deben agrupar los datos por UBIGEO y grupo de edad adulta mayor (60-79 y 80+) y no PAM (menor o igual a 59 años) para calcular la proporción de atenciones. Así se ha calculado y filtrado la cantidad de PAM por UBIGEO. Se presenta el dato referencial de Lima, Lima, Pueblo Libre con las siguientes cifras de atenciones en salud mental.

De acuerdo con las tablas anteriores, se señala como ejemplo, el porcentaje de atenciones en salud para PAM fue de 25.96% y de población no PAM de 74.04% para el UBIGEO 150121 (Lima, Lima, Pueblo Libre). De esta manera puede aplicarse un factor a las cadenas presupuestales de gasto corriente relativas a la salud mental usando este criterio de prorrateo.

Respecto al gasto en personas con discapacidades también se encuentra el portal de datos abiertos del Gobierno del Perú, *datasets* de las personas con alguna discapacidad inscritas en el Consejo Nacional para la Integración de la Persona con Discapacidad (CONADIS) que presentan información del UBIGEO y de rango de edades. El cálculo sigue el mismo criterio que en los casos anteriores. La siguiente tabla muestra también que se cuenta con UBIGEO y datos de edad en cada registro.

Tabla 9

Muestra del Registro Nacional de Personas con Discapacidad en Cajamarca

				RANGOS DE EDAD				
UBIGEO	PROVINCIA	DISTRITO	TOTAL DE PERSONAS	0-11	12-17	18-29	30-59	60+
61307	SANTA CRUZ	PULAN	1	0	0	1	0	0
61306	SANTA CRUZ	NINABAMBA	64	9	3	8	27	17
61305	SANTA CRUZ	LA ESPERANZA	1	0	0	0	1	0
61304	SANTA CRUZ	CHANCAYBAÑOS	3	0	1	0	2	0
61302	SANTA CRUZ	ANDABAMBA	24	0	3	4	5	12
61301	SANTA CRUZ	SANTA CRUZ	1	0	0	1	0	0

Nota: Datos extraídos del Registro Nacional de Personas con Discapacidad (datosabiertos.gob.pe/dataset/mimp-personas-con-discapacidad-inscritas).

Con los resultados de los 4 *datasets* procesados se efectúa un proceso iterativo que toma todos los registros presupuestales no usados en los pasos previos y se procede con la asociación y el prorrateo de importes del Presupuesto Institucional Modificado (PIM) y del monto Devengado de cada cadena de la siguiente forma:

- Para la relación con el dataset de SUSALUD se filtran los registros presupuestales la función 20 (Salud) y la categoría de gasto 5 (Gasto corriente). Luego se ejecuta un script Python iterativo que busca el UBIGEO distrital de cada cadena y calcula y asigna el valor prorrateado a los nuevos campos: PIM_CALCULADO y DEVENGADO_CALCULADO.
- Para la relación con el dataset de SALUD mental se filtran los registros presupuestales la función 20 (Salud) y la categoría de gasto 5 (Gasto corriente). Luego se ejecuta un script iterativo que busca el UBIGEO a nivel de distrito de cada cadena y calcula y asigna el valor prorrateado a los nuevos campos: PIM_CALCULADO y DEVENGADO_CALCULADO considerando solo los códigos relativos a atenciones por diagnóstico de salud mental.
- Para la relación con el dataset del INPE se filtran los registros presupuestales la función 06 (Justicia), la categoría de gasto 5 (Gasto corriente) y el Pliego 061 que corresponde al INPE. Luego se ejecuta un script iterativo que busca el UBIGEO a nivel de distrito de cada cadena y calcula y asigna el valor prorrateado a los campos PIM y DEVENGADO calculados.
- Finalmente, para la relación con el dataset del CONADIS se filtran los registros presupuestales por la categoría de gasto 5 (Gasto corriente) y el Pliego 345 que corresponde al CONADIS. Luego se ejecuta un script que, en base al UBIGEO a nivel de distrito de cada cadena, calcula y asigna el valor prorrateado a los campos PIM y DEVENGADO calculados.

Finalmente, los registros involucrados en las relaciones señaladas son marcados como usados y se crea un dataset de cadenas no evaluadas para su envío al paso final de identificación por IA.

Identificaciones de gasto PAM con inteligencia artificial

La clasificación de cadenas del conjunto de datos residual se envía a evaluación por la IA Llama3 y solicita la identificación de cada cadena para ser clasificada como NO PAM, SI PAM y POSIBLE PAM.

Con la base filtrada con una gran cantidad de cadenas (estimadas en 600,000 por filtros manuales aplicados sobre la base de las cadenas de Protección y Asistencia Social) con potencial de ser PAM como de empleo, recreación, asistencia social, vulnerabilidad, etc., se lanza el *prompt* para la IA con un *script* de programa Python que debe permitir que la IA entienda el beneficio directo o indirecto a la PAM. El *prompt* propuesto es el siguiente:

Analiza estas las cadenas presupuestales peruanas que se envían para determinar si corresponden a gastos públicos orientados para adultos mayores (60 años a más).

Sigue las siguientes reglas para categorizar:

- SI PAM: Si prioriza directa o exclusivamente a Adultos Mayores (Por ejemplo: ancianos, pensiones de jubilación, gastos en personas cesantes, geriatría, cuidadores).
- POSIBLE PAM: temas relativos a empleo, oportunidades para adultos mayores, participación, salud o vulnerabilidad amplia (CANCER, CATARATAS EN LA VISTA, VACUNACIÓN, PASAMANOS, ENFERMEDADES DEGENERATIVAS, ALBERGUES, etc).
- NO PAM: Sin relación (sanidad animal, minería, exportaciones, agropecuaria, neonatos, educación infantil, energía, etc.).

Responde solo con un objeto de datos plano en texto que debe incluir ID numérico provisto de cada cadena y los valores de la clasificación: 'SI PAM', 'POSIBLE PAM' o 'NO PAM'. No agregar texto adicional, ni bloques de código, ni comentarios o explicaciones.

Ejemplo de salida esperada: {"89710": "SI PAM", "18981": "NO PAM"} }

Terminado el proceso, se procede con en los niveles anteriores a la separación de *dataset* de registros identificados y marca de cadenas usadas. Nótese en el ejemplo que el código “89710” o “18981” corresponde al ID de las cadenas presupuestarias clasificadas por la IA. Finalmente, se ejecuta un *script* de programa para asignar el identificador correspondiente a la temática hallada de acuerdo con el clasificador CLAGPAM que se muestra en la tabla.

Hallazgos derivados del análisis

Invisibilidad del gasto público para PAM

El análisis del presupuesto público (2019-2023) revela la coexistencia de estructuras de gasto que, si bien permiten identificar inversiones específicas categorizadas, presentan una fragmentación diagnóstica que invisibiliza el gasto integral en la población adulta mayor (PAM).

Limitaciones del clasificador funcional (CF)

Al analizar el gasto fuera del PpR mediante la clasificación funcional, el análisis de los datos presupuestales halló una nula visibilidad en los niveles de División Funcional y Grupo Funcional. La etiqueta “Adulto Mayor” o “Anciano” u otros términos absolutos para la PAM es inexistente en estos niveles jerárquicos de planificación presupuestal.

Con la identificación directa solo es posible al llegar al nivel de componente de gasto, específicamente bajo el código 5000451 - *Apoyo al Anciano*. No obstante, el análisis de los 5.2 millones de registros evidencia la marginalidad de este clasificador.

Tabla 10

Identificación de coincidencias de términos clave absolutos en cadenas presupuestales

NIVEL DE CLASIFICACIÓN	COINCIDENCIAS DIRECTAS (N)	PORCENTAJE DE IDENTIFICACIÓN
División Funcional	0	0.00%
Grupo Funcional	0	0.00%
Componente (Actividad/Proyecto)	1,038	0.02%
Producto/Proyecto no PpR	468	0.01%

Nota. Elaboración propia en base a cálculo por búsqueda términos clave PAM.

Brecha de cobertura temática frente a la Ley 30490

Al analizar y comparar los derechos reconocidos en la Ley PAM con la estructura presupuestal, se determinó que la cobertura temática directa es de apenas el 22.27%. Mientras que dimensiones como la seguridad social y los servicios especializados tienen una trazabilidad del 100%, dimensiones importantes como salud integral, trabajo, educación y accesibilidad no presentan ninguna trazabilidad directa en los clasificadores.

Esta desalineación normativa-presupuestal confirma que los clasificadores del presupuesto actual no están diseñados para monitorear el cumplimiento integral de la Ley 30490, limitándose a un enfoque que omite el gasto transversal ejecutado en otros sectores.

Limitaciones estructurales para la identificación integral del gasto PAM

El análisis de la estructura presupuestaria entre 2019 y 2023 permite establecer que la arquitectura actual para el seguimiento o monitoreo del gasto no está alineada con el enfoque de derechos de la Ley Nro. 30490. Se han determinado tres limitaciones críticas que impiden una cuantificación precisa.

1. Desarticulación entre los servicios CIAM y la estructura de programas

Aunque la Ley PAM establece que los gobiernos locales deben brindar servicios transversales (salud, educación, justicia, emprendimiento) a través de los Centros Integrales de Atención al Adulto Mayor (CIAM), el presupuesto solo captura estos esfuerzos de manera atomizada. Se observó que proyectos destinados a un mismo fin (ej. mejoramiento de un CIAM) poseen códigos y nombres específicos en los campos identificadores por distrito (ej. Lince, Wanchaq, Carabayllo), lo que impide una agregación directa del gasto en infraestructura para la PAM.

2. Ambigüedad en las categorías de “Población Vulnerable”

Una limitación severa es la falta de etiquetas por grupo de edad en las cadenas de gasto compartido. Al analizar componentes como el *Apoyo al ciudadano y familia en situación de vulnerabilidad social*, se identificaron 1,510 cadenas presupuestales donde el gasto para la PAM está diluido con otros grupos etarios.

- Solo el 0.25% de las cadenas en funciones clave (Salud, Protección Social, Trabajo) contienen términos explícitos (Adulto Mayor, Anciano, etc.).
- El 12.64% (553,223 cadenas) se dirige a “población vulnerable” en general, sin que exista un clasificador por edades que permita discernir la cuota de participación de la PAM en servicios de salud o justicia.

Síntesis de restricciones del modelo vigente

Tras el análisis de las cadenas presupuestales de funciones relativas a la PAM, se observa que la identificación del gasto es poco viable bajo la estructura presupuestal actual debido a:

Exclusividad temática rígida: El sistema presupuestal asigna importes a un solo tema, impidiendo registrar la naturaleza multisectorial del apoyo a la PAM.

Invisibilidad etaria: La ausencia de campos clasificadores por segmentos de edad imposibilita la trazabilidad en funciones no asistenciales (como educación o transporte).

Inconsistencia en el PpR: Los productos y proyectos del Presupuesto por Resultados no mantienen una nomenclatura estándar, dificultando la consolidación de datos para la toma de decisiones.

En resumen, la capacidad de identificación directa del gasto mediante los clasificadores vigentes es de solo el 22.27%. El 99.75% de las cadenas en funciones relevantes para la PAM carecen de términos específicos, lo que valida la necesidad de implementar un modelo de búsqueda basado en el diccionario de la Ley PAM y distribuciones proporcionales por uso de servicios de la PAM.

Algunos hallazgos de cadenas presupuestales identificadas

Aunque no es el objetivo del presente estudio, analizar las cifras de gasto público PAM, sino presentar un modelo de extracción de datos, se presentan ejemplos de cadenas identificables por el procedimiento.

Tabla 11

Ejemplo Nro. 1 de cadena presupuestal identificable para el gasto en la Población Adulta Mayor (PAM) de la Región Puno

DIMENSIÓN	VARIABLE	VALOR
Temporal y administrativa	Año	2020
	Nivel de gobierno	Gobierno Regional
	Sector	Gobiernos Regionales
	Pliego	Gobierno Regional del Dpto. de Puno
	Unidad ejecutora	Región Puno Sede Central
Estructura programática	Programa presupuestal	Asignaciones presupuestales que no resultan en productos
	Producto / Proyecto	Sin producto
	Componente actividad / Obra	Desarrollo social regional
Clasificación funcional	Función	Planeamiento, gestión y reserva de contingencia
	División funcional	Planeamiento gubernamental
	Grupo funcional	Planeamiento institucional
	Finalidad del gasto	Asilo de ancianos Virgen del Rosario de Chucuito
Ubicación geográfica	Región	Puno
	Provincia	Puno
	Distrito	Puno
Clasificación económica	Fuente de gasto	Recursos ordinarios
	Categoría de gasto	Gastos corrientes
	Genérica de gasto	Bienes y servicios
Montos	Monto presupuesto (S/)	34,675.00
	Monto ejecutado (S/)	34,675.00
Identificadores asociados con la PAM	CLAGPAM	10.1 (Vida digna, tabla 2)
	ID cadena presupuestal (en la base de datos)	488974
	Monto presupuesto PAM (S/)	34,675.00
	Monto ejecutado PAM (S/)	34,675.00

Nota. Elaboración propia a partir de los datos vinculados del Presupuesto Público y SUSALUD. El registro corresponde a una asignación presupuestal ejecutada en el departamento de Puno.

Tabla 12

Ejemplo Nro. 2 de cadena presupuestal identificable para el gasto en la población adulta mayor (PAM) de la Provincia Constitucional del Callao

DIMENSIÓN	VARIABLE	VALOR
Temporal y administrativa	Año	2021
	Nivel de gobierno	Gobierno Regional
	Sector	Gobiernos Regionales
	Pliego	Gobierno Regional de la PC del Callao
	Unidad ejecutora	Región Callao
Estructura programática	Programa presupuestal	Asignaciones presupuestales que no resultan en productos
	Producto / Proyecto	Sin producto
	Componente actividad / Obra	Apoyo a la persona adulta mayor
Clasificación funcional	Función	Protección social
	División funcional	Asistencia social
	Grupo funcional	Protección de poblaciones en riesgo
	Finalidad del gasto	Asistencia al anciano

Ubicación geográfica	Región	PC del Callao
	Provincia	Callao
	Distrito	Multidistrital
Clasificación económica	Fuente de gasto	Recursos determinados
	Categoría de gasto	Gastos corrientes
	Genérica de gasto	Bienes y servicios
Montos	Monto presupuesto (S/)	138,000.00
	Monto ejecutado (S/)	138,000.00
Identificadores PAM	CLAGPAM	10.1 (Vida digna, tabla 2)
	ID cadena presupuestal en la base de datos	0532912
	Monto presupuesto PAM (S/)	138,000.00
	Monto ejecutado PAM (S/)	138,000.00

Nota. Elaboración propia a partir de los datos vinculados del Presupuesto Público y SUSALUD. El registro corresponde a una asignación presupuestal ejecutada en la Provincia Constitucional del Callao.

DISCUSIÓN

El hallazgo principal de esta investigación es la existencia de una brecha crítica entre la ejecución presupuestal real y la visibilidad presupuestaria del gasto en la población adulta mayor en el Perú. Mientras que el método de cuantificación directa (basado solo en clasificadores presupuestarios actuales del MEF) identifica un promedio anual de S/ 8,540 millones, la potencial aplicación del modelo procedimental propuesto indica que esta cifra es bastante parcial dados los hallazgos muestrales de cadenas presupuestales adicionales de gasto PAM directo e indirecto.

Los resultados del Nivel 2 (identificación por clasificadores vigentes) revelaron que el 99.61% del gasto visible se concentra en programas de transferencia monetaria (Pensión 65). Sin embargo, al aplicar los niveles 3 y 4 se logra recuperar en mayor medida la trazabilidad del gasto en temas transversales que actualmente se reportan como «gastos generales», donde el componente más fuerte son los datos de servicios de salud para la PAM identificados y prorrateados. Esto coincide con los postulados de De La Fuente (2017): el gasto público debe prorratearse según el uso real de los servicios.

Eficacia vs. visibilidad

Es paradójico que, aunque el Estado peruano muestra un nivel de ejecución promedio «Bueno» (93.1%) en el gasto para la PAM para el periodo de estudio, la cobertura temática identificable de los derechos de la Ley 30490 sea de apenas el 22.27%. Esta investigación sostiene que la alta eficacia de ejecución reportada por el Presupuesto Público en favor de la PAM no es real, ya que solo mide lo que es fácilmente etiquetable. El modelo propuesto identifica que el 99.75% de las cadenas en funciones clave carecen de etiquetas etarias, lo que sugiere que el Estado gasta, pero no sabe con precisión *en quién* gasta.

CONCLUSIONES Y RECOMENDACIONES

Invisibilidad estructural. Los clasificadores presupuestarios vigentes en Perú (PpR y Funcional) son insuficientes para monitorear el cumplimiento de la Ley Nro. 30490. La ausencia de campos de segmentación etaria en las cadenas de gasto compartido genera una subestimación sistemática de la inversión pública en la PAM.

Potencial de los datos abiertos. La integración de fuentes de datos externas (SUSALUD, INPE, MINSA y CONADIS) mediante un modelo de proporcionalidad geográfica y demográfica permite «descubrir» gasto público que actualmente es invisible. Sin embargo, estos portales de datos abiertos no siempre usan formatos estándar en los repositorios de datos. La marcada heterogeneidad en los formatos de los *datasets* implica, como se ha descrito en el presente estudio, la necesidad de homogeneización y estandarización de estas estructuras para viabilizar el procesamiento.

Optimización de la gestión. El modelo procedimental y el diccionario de términos propuestos constituyen una herramienta de auditoría social y gestión técnica. Su implementación permitiría a los tomadores de decisiones reorientar recursos hacia áreas críticas como la salud geriátrica y la accesibilidad, superando el actual enfoque casi exclusivo en pensiones y algunos programas de cuidado del adulto mayor plasmados en el PpR.

Aplicaciones del estudio. El modelo propuesto ofrece un proceso metodológico sistemático que puede aplicarse a la identificación y cuantificación de datos de otros temas transversales en el presupuesto público como: otros

grupos poblacionales vulnerables (neonatos, niñez y adolescencia), gasto en desarrollo tecnológico, gasto en personas con discapacidades, interculturalidad, entre otros.

REFERENCIAS

- Banco Interamericano de Desarrollo - BID. (2020). *Guía de etiquetado y trazabilidad del gasto público social en América Latina*. Publicaciones BID.
- Chen, L., Zaharia, M., & Zou, J. (2023). FrugalGPT: How to use large language models while reducing cost and improving performance. arXiv. <https://arxiv.org/pdf/2305.05176>
- Fondo de las Naciones Unidas para la Infancia. (2020). *Actualización de la metodología para estimar el gasto público en niñas, niños y adolescentes*. UNICEF. <https://www.unicef.org/peru/informes/actualizacion-de-la-metodologia-para-estimar-el-gasto-publico-en-ninas-ninos-y-adolescentes>
- Grimmer, J., & Stewart, B. M. (2013). *Text as data: The promise and pitfalls of automatic content analysis methods for political texts*. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
- Hernández-Sampieri, R., Fernández-Collado, C., Baptista-Lucio, P., & Mendoza-Torres, C. P. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. Mc Graw Hill.
- Instituto Nacional Penitenciario. (2024). *Portal de datos abiertos del INPE*. <https://portal-inpe.opendata.arcgis.com/datasets/031da7cad9484b3a9223e6057f6180f8/about>
- Li, J., & Li, Y. (2024). A Sentence-Matching Model Based on Multi-Granularity Contextual Key Semantic Interaction. *Applied Sciences*, 14(12), 5197. <https://doi.org/10.3390/app14125197>
- Macanovic, A. (2022). Text mining for social science – The state and the future of computational text analysis in sociology. *Social Science Research*, 107, 102784. <https://www.sciencedirect.com/science/article/pii/S0049089X22000904>
- Ministerio de Economía y Finanzas - MEF. (2021). *Directiva para la Programación Multianual Presupuestaria y Formulación Presupuestaria*. Dirección General de Presupuesto Público, Lima, Perú.
- Ministerio de Economía y Finanzas. (2022). *Directiva N.º 0003-2022-EF/50.01: Directiva de Seguimiento en el marco de la Fase de Evaluación Presupuestaria*. <https://www.gob.pe/institucion/mef/normas-legales/3304740-0003-2022-ef-50-01>
- Ministerio de Economía y Finanzas. (2025). *Estructura programática de programas presupuestales: Combinación de cadenas funcionales 2025*. Dirección General de Presupuesto Público. <https://www.mef.gob.pe/es/presupuesto-por-resultados/instrumentos/programas-presupuestales>
- Ministerio de Economía y Finanzas. (2025). *Presupuesto y Ejecución de Gasto [Conjunto de datos]*. Portal de Datos Abiertos del MEF. <https://datosabiertos.mef.gob.pe/dataset/presupuesto-y-ejecucion-de-gasto>
- Ministerio de la Mujer y Poblaciones Vulnerables [MIMP]. (2021). *Política Nacional Multisectorial para las Personas Adultas Mayores al 2030* (Decreto Supremo N° 006-2021-MIMP). Gobierno del Perú.
- Ministerio de la Mujer y Poblaciones Vulnerables. (2016, 21 de julio). *Ley N° 30490: Ley de la Persona Adulta Mayor*. Diario Oficial El Peruano. <https://www.gob.pe/mimp>
- Ministerio de Salud. (s.f.). MINSA. *Datos de salud mental [Conjunto de datos]*. Plataforma Nacional de Datos Abiertos. <https://www.datosabiertos.gob.pe/dataset/minsa-salud-mental/>
- Ministerio del Ambiente. (2015). *Caracterización y cuantificación del gasto público ambiental peruano* (Equipo técnico: A. Urriola Manrique, F. León Shedán y N. Shack Yalta). Plataforma Digital Única del Estado Peruano (gob.pe). <https://www.gob.pe/institucion/minam/informes-publicaciones/2679-caracterizacion-y-cuantificacion-del-gasto-publico-ambiental-peruano>
- Navarrete Talavera, J. F. (2024). *Perú: Modelo procedimental de gestión pública presupuestal para mejorar la cuantificación y análisis del gasto en personas adultas mayores (2019-2023)* [Tesis de maestría, Universidad San Martín de Porres].
- Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems with Applications*, 42(9), 4348–4360. <https://doi.org/10.1016/j.eswa.2015.01.050>
- Oficina Nacional de Pensiones. (2024). *Sistema Previsional Peruano*. https://www.onp.gob.pe/pensiones_peru_onp/sistema_previsional
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Reveilhac, M., & Morselli, D. (2022). Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data. *Political Research Exchange*, 4(1). <https://doi.org/10.1080/2474736X.2022.2029217>
- Shack, N. (2014). *Gasto público ambiental: Los casos del Perú y El Salvador* (Serie Macroeconomía del Desarrollo N.º 158; LC/L.3913). Comisión Económica para América Latina y el Caribe.
- Superintendencia Nacional de Salud. (2024). *Portal de datos abiertos de SUSALUD [Conjunto de datos]*. <http://datos.susalud.gob.pe/dataset>.