

EQUIDAD DE INGRESOS POR PENSIONES EN EL SISTEMA PREVISIONAL PERUANO: PERIODO 2012-2022

INCOME EQUITY IN THE PERUVIAN PENSION SYSTEM: PERIOD 2012-2022

Recibido: 18 de marzo de 2024 | Aprobado: 9 de julio de 2024

MARCO AURELIO ESPINOZA RIVERA¹

MESPINOZA.PE@GMAIL.COM

ORCID: 0000-0002-5089-938X

ISAAC ROBERTO ÁNGELES LAZO²

IANGELESL@USMP.PE

ORCID: 0000-0003-4224-6415

RESUMEN

Este estudio profundiza en el sistema previsional peruano entre 2012 y 2022, con un énfasis particular en la equidad de los ingresos por pensiones de jubilación y los factores determinantes de estas inequidades. A través de un enfoque cuantitativo y la implementación de técnicas avanzadas de machine learning, incluyendo modelos de regresión XGBoost y análisis SHAP, se analizó una muestra representativa de 858,474 pensionistas. Los resultados revelan una discrepancia significativa entre los pensionistas del Sistema Nacional de Pensiones y aquellos del Sistema Privado de Pensiones, siendo las pensiones de este último, en promedio, más elevadas.

El estudio identifica como factores cruciales en la equidad de ingresos por pensiones el nivel educativo, la ubicación geográfica (Lima versus otras regiones), la edad, el género y el sector de empleo (público o privado). Sorprendentemente, el nivel educativo de los pensionistas influye en casi un 20% del monto de la pensión por jubilación, lo que subraya su papel significativo en la determinación de los ingresos de jubilación. Este hallazgo es de particular importancia para la formulación de políticas públicas, sugiriendo que mejorar el acceso a la educación podría ser una estrategia clave para aumentar la equidad en el sistema previsional peruano.

Palabras clave: Sistema previsional, Inequidad de ingresos por pensión, Igualdad y justicia en pensiones.

ABSTRACT

This study delves into the Peruvian pension system between 2012 and 2022, with a particular emphasis on the equity of retirement pension incomes and the determinants of these inequities. Employing a quantitative approach and the implementation of advanced machine learning techniques, including XGBoost regression models and SHAP analysis, a representative sample of 858,474

pensioners was analyzed. The findings reveal a significant discrepancy between pensioners in the National Pension System and those in the Private Pension System, with the latter's pensions being on average higher.

The study identifies key factors in pension income equity as educational level, geographic location (Lima versus other regions), age, gender, and employment sector (public or private). Notably, the educational level of pensioners accounts for almost 20% of the retirement pension amount, highlighting its significant role in determining retirement income. This finding is of particular importance for public policy formulation, suggesting that improving access to education could be a key strategy to enhance equity in the Peruvian pension system.

Keywords: Pension system, Pension income inequality, Equality and justice in pensions.

1 Doctor en Gobierno y Política Pública.
Instituto de Gobierno y de Gestión Pública.

2 Doctor en Sociología; Universidad
Nacional Mayor de San Marcos.

INTRODUCCIÓN

A fines del año 2022, una de cada nueve personas en el mundo tenía más de 60 años y este número seguirá aumentando si la gente sigue teniendo menos hijos y vive más años; esto significa que en un futuro cercano habrá cada vez más personas mayores en la población global (Fondo de Población de las Naciones Unidas, 2022). Este aumento en la población ha ocasionado implicaciones significativas en sistemas previsionales, al no haber considerado esta nueva realidad demográfica.

Vivir más tiempo no es malo; de hecho, es algo que la humanidad ha querido lograr desde hace tiempo, la ciencia está haciendo un gran trabajo para alcanzar esa meta. Sin embargo, este cambio trajo desafíos para la economía y la sociedad, especialmente en países como el nuestro donde hay mucha desigualdad. Este problema podría empeorar para los grupos más frágiles, como las personas mayores. Hay planes y acuerdos internacionales, como el *Plan de acción de Madrid* y la *Convención Interamericana de 2015*, que buscan proteger los derechos y mejorar la calidad de vida de las personas mayores. Estos acuerdos nos guían sobre cómo crear un entorno más amigable para el envejecimiento (Organismo de Estados Americanos, 2022). La prolongación de la esperanza de vida, aunque sea un logro humano, plantea retos en la gestión pública y en sistemas como el previsional; por lo que requieren una acción coordinada y fundamentada en el respeto de los derechos humanos.

En esta línea de crear entornos favorables para el envejecimiento, el investigador consideró pertinente posicionar el tema del pensionamiento con equidad, para su discusión en la política pública del país. El Perú empezó su sistema de pensiones a mediados del siglo pasado con un sistema basado en contribuciones colectivas (SNP), que luego de algunos años empezó a tener problemas financieros. Para solucionarlos, en 1992 se creó un nuevo sistema basado en cuentas individuales de ahorro (SPP), que empezó a funcionar al lado del antiguo sistema desde junio de 1993. Al momento de la investigación, los trabajadores tienen que elegir entre estos dos sistemas para hacer sus aportes obligatorios, pese a que existen otros 16 más pequeños sistemas que en total cubren alrededor del 2.4% de la fuerza laboral (Ministerio de Economía y Finanzas, 2004). Esto crea un panorama complejo que la administración pública debe afrontar con estrategias y políticas innovadoras.

A junio de 2022 (Asociación de AFP, 2022), existían 4.706 millones de afiliados al SNP, de los cuales 1.459 millones (31.0%) son cotizantes activos. La cotización obligatoria es del 13% de la remuneración que perciben. El SNP tiene 660,998 pensionistas cuya pensión promedio es de S/ 767.40 soles y que representan el 77.0% de los pensionistas del Sistema Previsional Peruano (Oficina de Normalización Previsional, 2022).

A la misma fecha, existían 8.478 millones de personas afiliadas al Sistema de cuentas individuales de ahorro (SPP), de los cuales, 3.561 millones (42.0%) son cotizantes activos o regulares. El aporte a su fondo individual de pensión es del 10% de la remuneración que perciben, pero también deben pagar otros costos como una tarifa de administración y un seguro que cubre invalidez y otros gastos, costos adicionales que hacen que el aporte sea casi el 13% de la remuneración mensual. A junio del 2022, el SPP tiene 197,476 pensionistas cuya pensión promedio es de S/ 987.00 soles y que representan el 23.0% de los pensionistas del Sistema Previsional Peruano (Superintendencia de Banca, Seguros y AFP, 2022).

Al momento de la presente investigación, el sistema previsional presenta rasgos de fragmentación e inequidad, dado que, dos personas con situaciones parecidas reciben pensiones muy distintas dependiendo si están en el sistema privado (SPP) o en el público (SNP), pudiendo resultar que las pensiones del sistema privado son bajas y en el sistema público podría quedarse sin pensión si no se ha contribuido por al menos 10 años. Este panorama muestra la importancia de entender bien las complejidades y diferencias entre los sistemas de pensiones disponibles.

En base a lo anteriormente expresado, el investigador consideró como objetivo de la investigación el analizar la equidad en la distribución de ingresos por pensiones en el sistema previsional peruano, identificando los factores clave que influyen las disparidades y evaluar cómo estos factores se correlacionan con las condiciones socioeconómicas y demográficas de los pensionistas, durante el periodo comprendido entre los años 2012 al 2022.

Asimismo, el autor declara como hipótesis general: La equidad en la distribución de ingresos por pensiones en el sistema previsional peruano entre 2012 y 2022 ha sido significativamente influenciada por factores socioeconómicos y demográficos específicos, resultando en disparidades notables. Consignando como hipótesis específicas, las siguientes:

- a. Los pensionistas del Sistema Previsional Peruano que residen en ciertas zonas geográficas reciben ingresos significativamente diferentes, contribuyendo a la disparidad en la distribución de ingresos.
- b. El tipo de entidad laboral (pública o privada) en la que trabajó el pensionista tiene un impacto considerable en la cantidad de sus ingresos de pensiones.

- c. Existe una correlación significativa entre el sistema previsional receptor de los aportes y la cantidad de ingresos que recibe el pensionista.
- d. El género del pensionista es un factor determinante en la disparidad de los ingresos por pensiones en el sistema previsional peruano.
- e. El nivel de educación alcanzado por los pensionistas afecta de manera significativa sus ingresos por pensiones.

El estudio aborda un tema crítico para la sociedad peruana, dada la importancia de los sistemas de pensiones en la seguridad económica y el bienestar de los ciudadanos mayores. La investigación pone de relieve las disparidades existentes en el sistema de pensiones, incluyendo diferencias basadas en la ubicación geográfica, género, nivel educativo, y entre los sistemas de pensiones público y privado. Los hallazgos proporcionan una base empírica para el diseño y la implementación de políticas públicas más equitativas y efectivas en el ámbito de las pensiones, contribuyendo así a un sistema previsional más justo y sostenible. Finalmente, los resultados son de gran valor para una variedad de *stakeholders*, incluyendo formuladores de políticas, administradores de sistemas de pensiones, organizaciones de la sociedad civil, y los propios pensionistas, al ofrecer datos e *insights* detallados sobre el estado actual del sistema.

DESARROLLO

A pesar de que en los últimos tiempos han surgido enfoques meta sobre la justicia, que buscan armonizar o conciliar diversas teorías de la justicia, como lo ilustra el caso de (Stilman, 2022), quien argumenta que una sociedad justa será aquella que logre amalgamar bajo el principio democrático, las concepciones particulares de la justicia sostenidas por sus miembros. En este estudio, el investigador sostiene que las teorías de justicia más influyentes en la actualidad son: la noción de “justicia como equidad” de John Rawls y la teoría de la “igualdad de capacidades” de Amartya Sen. Rawls formula una teoría distributiva de la justicia que sigue líneas liberales, buscando una aceptación universal y otorgando prioridad a la libertad sobre la igualdad. No obstante, contempla la necesidad de corregir las desigualdades en la libertad mediante una apuesta por la igualdad de oportunidades. Amartya Sen, por su parte, dirige el debate sobre la igualdad no hacia la pregunta de “¿por qué igualdad?”, sino más bien hacia la interrogante de “¿igualdad de qué?”. Para él, todas las teorías sobre la estructura social tienden a exigir un tratamiento equitativo de los individuos en alguna dimensión relevante, en relación a alguna variable crucial para la teoría en cuestión.

Para (Sen, 1999) la equidad no se limita a la igualdad formal, sino que también incluye la igualdad de oportunidades y el acceso a recursos básicos; la equidad se alcanza cuando todas las personas tienen la capacidad de tomar decisiones y llevar a cabo sus propias vidas de acuerdo con sus propios valores y objetivos. En contraste, (Nussbaum, *Crear capacidades: Propuesta para el desarrollo humano*, 2012) sostiene que la equidad significa tratar a todas las personas como seres humanos dignos de respeto y consideración, y se enfoca en promover y proteger las capacidades humanas básicas como la salud, la educación y la libertad de expresión. (Rawls, *Teoría de la Justicia*, 2012) por su parte, propone una teoría de la justicia que se basa en el concepto de equidad y establece que, las desigualdades sociales y económicas solo son justas si benefician a las personas menos favorecidas de la sociedad.

El enfoque de investigación es de tipo cuantitativo, de diseño no experimental, transversal; descriptivo y analítico. Se utilizaron datos secundarios de la Encuesta Nacional de Hogares (ENAHOG). La población de estudio estuvo conformada por todos aquellos que a junio de 2022 son pensionistas del sistema previsional peruano: 858,474 pensionistas (660,998 pertenecen al SNP, y 197,476 al SPP (Asociación de AFP, 2022).

En el contexto de la metodología de investigación científica, el uso de bases de datos estructurados como fuente de datos como alternativa al trabajo de campo, es una tendencia que ha ido ganando aceptación y reconocimiento en el ámbito académico y de la investigación. Uno de los autores más destacados en esta área es Yin, quien argumenta que las bases de datos estructurados pueden ser una fuente de datos válida y confiable para la investigación científica, y pueden ser utilizadas para realizar una amplia gama de investigaciones, incluyendo estudios de caso, estudios transversales y estudios longitudinales (Yin, 2014). Quinn Patton, señala que las bases de datos estructurados pueden ser una herramienta valiosa para la recopilación y organización eficiente y efectiva de datos cualitativos (Quinn Patton, 2015). Collier, señala que las bases de datos estructurados son una herramienta esencial para la gestión de datos de investigación, puesto que pueden ayudar a los investigadores a organizar, almacenar, preservar y compartir sus datos de manera segura y eficiente (Collier, 2016).

En esta investigación, al recurrir a bases de datos estructurada con millones de registros, se necesita contar con herramientas y técnicas eficientes que permitan extraer los datos de manera fiable y efectiva. Según (McKinney, 2017), Python se ha convertido en una opción destacada para la recolección de datos desde bases de datos estructuradas con millones de registros, debido a su capacidad de conectarse a estas bases de datos y a la disponibilidad de gran variedad de bibliotecas especializadas (*Deep learning*, *Machine learning*, Cálculo numérico, Visualización, Inteligencia artificial explicable, Procesamiento de lenguaje natural).

Una vez que los datos han sido recolectados, se utilizaron diversas técnicas de procesamiento para el análisis de la información, entre ellas:

- a. *Limpieza de datos*: Se identificaron y corrigieron problemas o errores en los datos recolectados, como valores atípicos, datos faltantes o inconsistentes. Se reemplazaron valores faltantes, y se detectaron y eliminaron valores atípicos.
- b. *Transformación de datos*: Se codificaron las variables categóricas y se crearon nuevas variables derivadas a partir de las existentes. Se generan variables *dummy* a las características: zona geográfica, entidad de trabajo, sistema pensionario, sexo.
- c. *Minería de datos*: Esta técnica implica descubrir patrones, relaciones y conocimientos útiles a partir de data ENAHO. Se utilizaron algoritmos de aprendizaje automático y técnicas estadísticas para explorar los datos y encontrar información oculta. Se emplearon como filtros básicos, los datos correspondientes a monto de pensión, y sistema previsional (AFP/ONP); y como características asociadas a los ingresos del pensionista aquellas que plantea el investigador en su hipótesis principal: zona geográfica (ubigeo), entidad donde trabajó la persona antes de su jubilación, género, edad, y nivel de estudio alcanzado.
- d. *Ingeniería de datos*: Se utilizó el *software* R 4.2.2 para la determinación del porcentaje de participación, ingresos para cada sistema de pensiones, en la ventana de tiempo conformada entre los años 2012 al 2022 con intervalo de 2 años, además del coeficiente Gini para determinar la equidad entre los sistemas pensionarios, la información fue segregada por zonas: costa, sierra, selva, Lima metropolitana y total Perú.
- e. *Modelado y predicción*: Se utilizó específicamente la técnica de árboles de regresión potenciados por gradientes XGBoost para modelar la relación entre la variable dependiente y las variables independientes. Finalmente, se trabajó con modelamientos de *Machine learning* por ser técnicas modernas con mayor capacidad para generar predicciones con cercanía a la realidad; además, de ser el estado del arte en analítica de datos, se propuso modelos de regresión de árboles de decisión.
- f. *Analítica de datos*: Para la evaluación de los factores de impacto se utilizó el diagrama de cascada (*waterfall*), y todo el proceso de analítica de datos fue llevado a cabo mediante el *software* Python 3.11.2 con la librería SHAP.

RESULTADOS

Teniendo en consideración que la hipótesis de investigación es “La equidad en la distribución de ingresos por pensiones en el sistema previsional peruano entre los años 2012 y 2022 ha sido significativamente influenciada por factores socioeconómicos y demográficos específicos, resultando en disparidades notables”, se describen en consistencia; los resultados obtenidos.

Ingreso por pensión, como factor de equidad

Para realizar este primer análisis de datos, se utilizó técnicas de minería de datos con la finalidad de ubicar indicadores que permitan conocer la evolución de la pobreza monetaria, del bienestar y de las condiciones de vida de los hogares, además de efectuar diagnósticos sobre las condiciones de vida y pobreza monetaria de la población. Se consideraron los datos relacionados a ingresos por pensiones de jubilación, segmentándolos asimismo por fondo pensionario (AFP, ONP). Finalmente, se identificaron las características que, según juicio experto del investigador, impactan en los ingresos de los pensionistas: zona geográfica (ubigeo), tipo de entidad donde trabajó la persona antes de su jubilación, sistema pensionario, sexo, edad, nivel de estudio e ingreso por pensión de jubilación en soles (S/).

Tabla 1. Ingreso mensual promedio, por fondo previsional, Perú, 2012 al 2022

Periodo	Ingresos (S/)		Participación (%)	
	AFP	ONP	AFP	ONP
2012	618.5	591.3	13.9	86.1
Lima Metropolitana	694.4	604.9	18.6	81.4
Región Costa	596.3	558.1	11.9	88.1
Región Sierra	614.7	634.1	10.9	89.1
Región Selva	568.7	567.9	20.8	79.2

EQUIDAD DE INGRESOS POR PENSIONES EN EL SISTEMA PREVISIONAL PERUANO: PERIODO 2012-2022

2014	725.0	608.3	14.8	85.2
Lima Metropolitana	1062.3	613.6	17.1	82.9
Región Costa	611.0	578.3	10.1	89.9
Región Sierra	654.2	628.3	15.3	84.7
Región Selva	572.6	612.9	41.3	58.7
2016	771.0	611.5	17.2	82.8
Lima Metropolitana	955.4	622.1	19.1	80.9
Región Costa	805.6	599.7	14.2	85.8
Región Sierra	710.9	592.0	16.2	83.8
Región Selva	612.2	632.3	39.8	60.2
2018	854.9	775.9	13.5	86.5
Lima Metropolitana	921.7	790.6	16.5	83.5
Región Costa	790.6	765.0	10.8	89.2
Región Sierra	853.5	763.5	12.4	87.6
Región Selva	854.0	784.4	21.1	78.9
2020	788.2	777.2	13.8	86.2
Lima Metropolitana	1014.8	774.8	16.1	83.9
Región Costa	745.1	769.4	11.6	88.4
Región Sierra	735.5	765.9	12.5	87.5
Región Selva	657.4	798.7	21.3	78.7
2021	883.4	786.7	13.2	86.8
Lima Metropolitana	1138.7	810.5	17.5	82.5
Región Costa	866.0	768.4	11.5	88.5
Región Sierra	739.3	781.8	9.1	90.2
Región Selva	789.7	786.1	22.9	77.1
2022	978.6	796.2	12.6	87.4
Lima Metropolitana	1262.6	846.2	18.9	81.1
Región Costa	986.9	767.4	11.4	88.6
Región Sierra	743.0	797.7	5.8	94.2
Región Selva	921.9	773.6	24.4	75.6

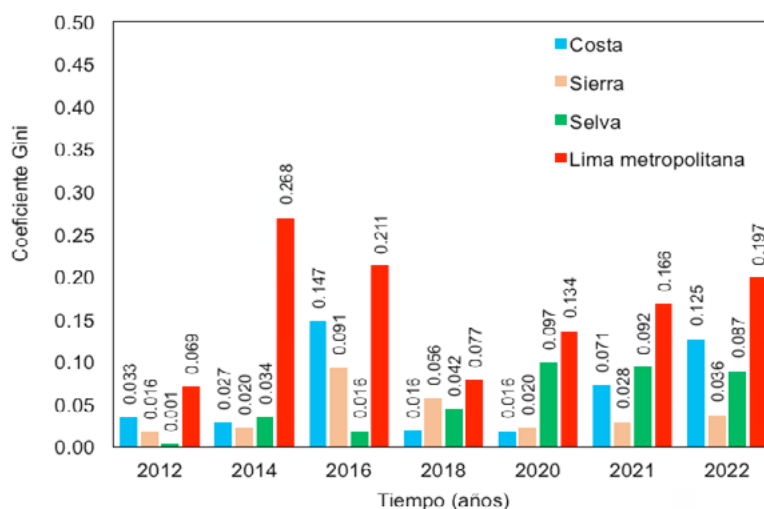
Fuente: INEI, Encuesta Nacional de Hogares Actualizada (ENAHO 2022)

La Tabla 1 muestra la media de ingresos por concepto de pensión tanto en el fondo de pensiones que administra el Estado peruano, como el que es administrado por el sector privado en el periodo comprendido entre los 2012 al 2022, donde se aprecia una tendencia creciente en ambos fondos, siendo más notorio en la AFP (valores estuvieron entre S/ 618.5 a S/ 978.6) a comparación de la ONP (valores estuvieron entre S/ 519.3 a S/ 796.2).

Coeficiente Gini de equidad de ingresos por pensión

El coeficiente de Gini es una medida de desigualdad de ingresos. El valor de este coeficiente varía entre 0 y 1, donde 0 refleja perfecta igualdad (todos tienen el mismo ingreso) y 1 refleja perfecta desigualdad (una persona tiene todo el ingreso y todos los demás no tienen ingreso).

Figura 1. Coeficiente de equidad Gini según Región, 2012 al 2022



Fuente: Elaboración propia

En la Figura 1 se observa que la desigualdad en Lima Metropolitana es consistentemente más alta que la desigualdad promedio en el país. Esto podría indicar que dentro de la capital hay grandes disparidades entre los pensionistas, posiblemente debido a la heterogeneidad del mercado laboral, las diferencias en oportunidades educativas y laborales, entre otros factores. La Selva, por otro lado, tuvo la desigualdad más baja en 2012, pero ha visto un aumento en los años posteriores, aunque sigue siendo relativamente baja. La Costa y Sierra han mostrado fluctuaciones, pero en general la Sierra ha mantenido un coeficiente Gini más bajo en comparación con la Costa, lo que sugiere una distribución más equitativa entre pensionistas en la sierra.

Tabla 2. Ingresos y coeficiente de equidad Gini según sistema pensionario, 2022

Región	Sistema de pensión	%	S/	Coeficiente de equidad Gini
Costa	AFP	11.4	986.9	0.125
	ONP	88.6	767.4	
Sierra	AFP	5.8	743.0	0.035
	ONP	94.2	797.7	
Selva	AFP	24.4	921.9	0.087
	ONP	75.6	773.6	
Lima Metropolitana	AFP	18.9	1262.6	0.197
	ONP	81.1	846.2	
Total, Perú	AFP	12.6	978.6	0.103
	ONP	87.4	796.2	

Fuente: Elaboración propia

En la Tabla 2 se puede observar que, en todas las regiones, la mayoría de los pensionistas están en el sistema público (ONP). Esto refleja la naturaleza del sistema pensional en Perú, donde el sistema público es más popular y accesible para la mayoría de la población. En cuanto a las pensiones promedio, el sistema privado (AFP) tiende a pagar pensiones más altas en todas las regiones. Sin embargo, hay que destacar que el sistema privado representa a un menor porcentaje de pensionistas en todas las regiones. Esto podría indicar que aquellos en el sistema privado tienen condiciones económicas previas más ventajosas o diferentes condiciones de aporte.

Lima Metropolitana presenta la brecha más grande entre ONP y AFP, con una diferencia de 416.4 soles. Además, es la región con la pensión promedio más alta en ambos sistemas. Esto puede reflejar el costo de vida más alto en la capital y/o la concentración de trabajadores con salarios más altos que se jubilan con pensiones más sustanciales. Sorprendentemente, en la Sierra, el sistema público (ONP) paga una pensión promedio más alta que el sistema privado (AFP). Esta es la única región donde esto ocurre. La razón podría ser la representación demográfica o

laboral particular de aquellos que están en el sistema privado en esta región, o puede estar relacionado con las particularidades del mercado laboral en la Sierra.

A nivel nacional, las cifras reflejan lo que se ve regionalmente: la mayoría están en el sistema público, pero las pensiones en el sistema privado son, en promedio, más altas. Esto evidencia desigualdad entre los pensionistas de ONP y AFP en cuanto a las cantidades de pensiones. Esto puede ser una señal de que hay inequidades en el acceso a oportunidades laborales remuneradas, en la capacidad de contribuir a un sistema privado y en las rentabilidades o beneficios del sistema privado.

Finalmente, si se busca ajustar los valores reales de ingresos promedio por pensión en función de su proporcionalidad, se debería ponderar los ingresos promedio de cada sistema (ONP y AFP) por la proporción de pensionistas en cada sistema, lo que resulta en S/ 820.35 soles.

Exploración de factores de equidad

Para este análisis se utilizó técnicas de ingeniería de datos sobre la data de la ENAHO, por lo cual se generaron variables *dummy* a las características: zona geográfica obteniendo nuevas variables de Costa, Sierra, Selva y Lima Metropolitana. Para la característica entidad de trabajo al ser una variable de escala nominal se consideró como 0=entidad pública y 1=entidad privada. Para la característica sistema pensionario se consideró 0=ONP y 1=AFP. Para la característica sexo se codificó 0=mujer y 1=hombre. Para la característica edad se encontró en escala numérica discreta. Y finalmente, para la característica nivel de estudio, una escala ordinal de 10 puntos (1=educación inicial, 2=primaria incompleta, 3=primaria completa, 4=secundaria incompleta, 5=secundaria completa, 6=superior no universitaria incompleta, 7=superior no universitaria completa, 8=superior universitaria incompleta, 9=superior universitaria completa, hasta 10=maestría/doctorado).

Los modelos de regresión de árboles de decisión son herramientas poderosas en el campo de la estadística y el análisis de datos, son una extensión de los árboles de decisión y generalmente se emplean para predecir valores numéricos en función de variables independientes. Capturan patrones no lineales y complejos en los datos, lo que los hace aptos para problemas en los que las relaciones entre variables son más intrincadas. Los árboles de regresión fueron propuestos por Leo Breiman en el libro *Classification and regression trees* (Breiman y otros, 1984) y son una técnica de aprendizaje supervisado que predice valores de respuestas mediante el aprendizaje de reglas de decisión derivadas de características.

Por otro lado, los algoritmos de *gradient boosting* son técnicas avanzadas de aprendizaje automático que se utilizan para construir modelos predictivos precisos y complejos, donde el objetivo es predecir valores numéricos o asignar etiquetas a diferentes categorías. La idea fundamental detrás de estos algoritmos es combinar múltiples modelos más simples, generalmente árboles de decisión, en un modelo más fuerte y preciso. El proceso se realiza en etapas iterativas, donde cada nuevo modelo se construye para corregir los errores del modelo anterior. Los modelos XGBoost, LightGBM y CatBoost son tres implementaciones avanzadas de algoritmos de *gradient boosting* que se han vuelto muy populares en la comunidad de ciencia de datos y análisis de datos. XGBoost (Extreme Gradient Boosting) es una librería optimizada y eficiente para *gradient boosting*, combina múltiples modelos más simples, generalmente árboles de decisión, y una de sus características distintivas es su capacidad para manejar tanto problemas de regresión como de clasificación; utiliza una función de pérdida personalizable, una función de regularización para controlar el sobreajuste, y maneja valores faltantes en los datos de manera efectiva. LightGBM (*Light Gradient Boosting Machine*), es otra librería de *gradient boosting* que se centra en la velocidad y la eficiencia; utiliza técnicas como “*binning*” (agrupación) y “*leaf-wise*” (crecimiento de hojas) para reducir el tiempo de entrenamiento y mejorar el uso de la memoria, también puede manejar tareas como ranking y predicción de tiempo de supervivencia. CatBoost (*Categorical Boosting*), es un algoritmo de *gradient boosting* que se destaca por su capacidad para manejar características categóricas de manera eficiente sin la necesidad de convertirlas en números; esto es especialmente útil en conjuntos de datos que contienen muchas variables categóricas, implementando regularización y ajustando automáticamente los hiperparámetros.

Tabla 3. Bondad de ajuste de modelos de regresión de árboles de decisión

Árboles de decisión potenciados por gradientes	Modelos	R ²
Extreme Gradient Boosting	XGBoost	0.897
Light gradient-boosting machine	LightGBM	0.470
Categorical Boosting	CatBoost	0.759

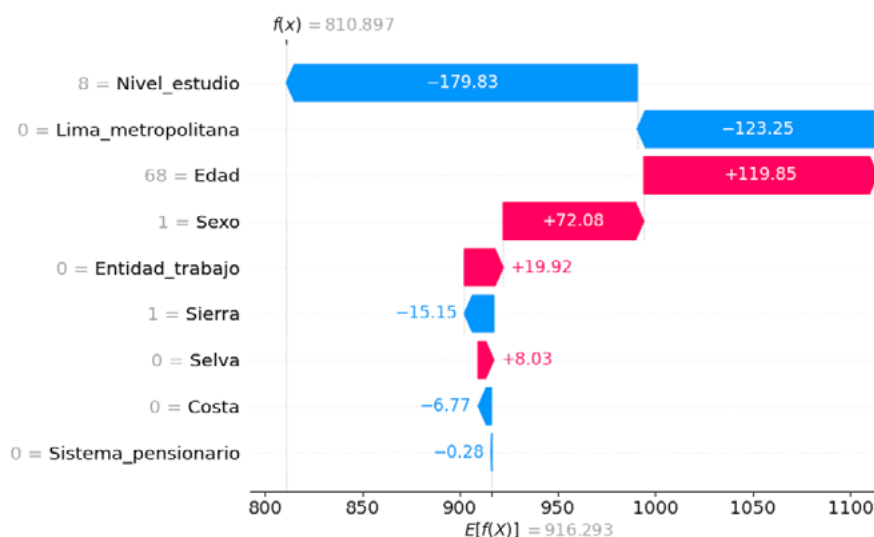
Fuente: Elaboración propia

R cuadrado (R^2), es una métrica estadística utilizada en el análisis de regresión para evaluar cuánta variabilidad en la variable dependiente puede ser explicada por las variables independientes, vale decir; mide qué tan bien el modelo de regresión se ajusta a los datos observados. Tiene valores que van desde 0 hasta 1, donde $R^2=0$ significa que el modelo no puede explicar nada de la variabilidad de los datos y todas las predicciones son equivalentes al valor promedio de la variable dependiente, y $R^2=1$ significa que el modelo explica toda la variabilidad de los datos y las predicciones coinciden perfectamente con los valores reales; consecuentemente, un R^2 más alto indica que el modelo es capaz de capturar y explicar una mayor cantidad de la variación en los datos observados. En la Tabla 3, se presenta la bondad de ajuste de modelos de regresión de árboles de decisión potenciados por gradientes para la evaluación de factores en la determinación del nivel de equidad de ingresos por pensiones otorgados por el sistema previsional peruano para el periodo 2022, donde se observa que el modelo con mayor coeficiente de determinación fue el Extreme Gradient Boosting (XGBoost) con R^2 de 0.897 (debiendo ser superior a 0.85 para ser considerado un modelo con buena bondad de ajuste), a comparación de LightGBM (0.470) y CatBoost (0.759).

Determinación de factores de mayor impacto sobre los ingresos

XGBoost, se sustenta en generar múltiples modelos secuenciales de predicción “sencilla”, y cada nuevo modelo toma el resultado del modelo anterior, haciendo que los resultados obtenidos en cada secuencia sean cada vez más “robustos”, y, por tanto, obteniendo en cada iteración resultados más exactos. Al aplicar este modelo mediante el software R v4.2.2, sobre la data publicada en la ENAHO, se obtuvieron los siguientes resultados:

Figura 2. Significancia de factores de equidad de ingresos por pensión



Fuente: Elaboración propia

Estos resultados indican que a menor nivel de estudio los ingresos por pensiones tienden a ser menores. Por ejemplo, los pensionistas con estudios superiores incompletos, tienen ingresos que superan en S/ 179.83 de aquellos que tienen estudios no superiores completos -escala inmediata anterior-. Esto se explica fundamentalmente porque las personas con un nivel de estudios más elevado suelen acceder a trabajos que requieren habilidades especializadas y conocimientos avanzados, por lo que suelen estar mejor remunerados. Dado que las personas con niveles de educación más altos tienen salarios más altos en promedio, sus contribuciones también serán más altas; como resultado, el cálculo de sus pensiones estará basado en un monto de contribuciones más sustancial, lo que podría resultar en una pensión más elevada. Por otro lado, las personas con niveles de educación más altos también pueden tener más oportunidades de ascenso y promoción en sus trabajos; esto puede llevar a un aumento gradual en sus ingresos a lo largo de su carrera, lo que a su vez influye en las contribuciones al sistema de pensiones y en la base de cálculo para su pensión futura. Finalmente, también se podría decir que los individuos con educación superior a menudo tienen acceso a trabajos más estables y de mayor calidad, lo que puede resultar en carreras laborales más prolongadas y contribuciones al sistema de pensiones durante más tiempo; contribuyendo a acumular un historial de contribuciones más amplio, lo que puede impactar positivamente en el monto de la pensión al momento de jubilarse.

En esta misma corrida de *machine learning*, aparece como segundo factor de importancia el lugar en el que se desempeñó laboralmente el pensionista (Lima metropolitana, resto del país) siendo que aquellos pensionistas

que trabajaron en el resto del país tienen ingresos inferiores en S/ 123.25 que aquellos que trabajaron en Lima Metropolitana. Respecto a este factor, se puede señalar que Lima Metropolitana suele tener un costo de vida más elevado en comparación con otras regiones del país; los gastos asociados con vivienda, alimentos, transporte y otros servicios suelen ser más altos en una gran ciudad como Lima, esto significa que los trabajadores que han laborado en Lima Metropolitana podrían haber enfrentado mayores gastos a lo largo de su vida laboral en comparación con aquellos que trabajaron en áreas con un menor costo de vida; sin embargo, las remuneraciones en Lima Metropolitana tienden a ser más altas para compensar el mayor costo de vida. Esto significa que, los pensionistas que trabajaron en Lima podrían experimentar una menor diferencia entre sus ingresos por pensión y los gastos necesarios para mantener su calidad de vida, en comparación con aquellos que trabajaron en áreas con un costo de vida más bajo.

Como tercer factor de importancia se encuentra la edad, con coeficiente estimado de 119.85 (68 años), esto quiere decir que por cada año de vida los ingresos por pensiones disminuyen en promedio S/ 119.85. En relación a estos resultados, se debe mencionar que la edad del pensionista puede influir en los ingresos por pensión debido a los gastos médicos adicionales que tienden a aumentar con la edad, los pensionistas mayores podrían enfrentar mayores costos de atención médica y una mayor vulnerabilidad financiera en comparación con los más jóvenes. Estos factores pueden afectar su capacidad para mantener su calidad de vida después de cubrir los gastos médicos y otros costos de vida, lo que podría requerir un ingreso por pensión más alto para garantizar su bienestar en la jubilación.

Como cuarto factor de importancia se encuentra el género del pensionista donde el coeficiente estimado fue de 72.08 (hombre), esto quiere decir que los hombres tienen en promedio S/ 72.08 más de ingresos por pensión que las mujeres. Respecto a esto, podemos mencionar que, en muchos contextos, existe una brecha salarial de género, lo que significa que las mujeres tienden a ganar menos que los hombres por el mismo trabajo o por trabajos de valor similar, esta brecha salarial se refleja en las contribuciones que las mujeres realizan al sistema de pensiones a lo largo de su vida laboral, lo que obviamente puede dar como resultado una base de cálculo más baja para sus pensiones. Por otro lado, las mujeres a menudo enfrentan interrupciones en sus carreras laborales debido a responsabilidades familiares, como el cuidado de los hijos o familiares mayores; estas interrupciones pueden resultar en períodos de ingresos más bajos o en la falta de contribuciones al sistema de pensiones durante ciertos periodos, afectando directamente el monto de la pensión. Así mismo; las mujeres tienden a tener una mayor esperanza de vida en comparación con los hombres, a medida que las mujeres viven más tiempo en la jubilación, es posible que necesiten una pensión más alta para cubrir los gastos durante un período más prolongado.

El quinto factor en importancia es el tipo de sector al que pertenece la organización en la que trabajó el pensionista, resultando el coeficiente estimado de 19.92 (entidad pública), esto indica que jubilarse trabajando en una entidad pública mejora los ingresos por pensiones en S/ 19.92. Se podría decir que, en algunos casos, el sector público puede ofrecer salarios y beneficios más atractivos en comparación con el sector privado, los empleados del sector público pueden tener acceso a salarios más estables, y mejores prestaciones. El sector público a menudo se caracteriza por una mayor estabilidad laboral en comparación con el sector privado, los empleados del sector público podrían tener carreras laborales más prolongadas y contribuir durante más tiempo al sistema de pensiones; esto podría resultar en un historial de contribuciones más amplio y, por lo tanto, en una base de cálculo más sólida para la pensión.

Los demás factores resultantes en la misma corrida de datos aplicando *machine learning* no fueron considerados importantes debido a los bajos coeficientes estimados (en valor absoluto inferiores a 16).

Corroboración de hipótesis de investigación

En esta investigación se declaró como hipótesis principal, que la equidad en la distribución de ingresos por pensiones en el sistema previsional peruano entre 2012 y 2022 ha sido significativamente influenciada por factores socioeconómicos y demográficos específicos, resultando en disparidades notables. Los factores socioeconómicos y demográficos se refieren a características de una población que pueden influir en su comportamiento, necesidades y dinámica. Los factores socioeconómicos considerados por el investigador son: el monto de pensión, tipo de entidad laboral (sector público o privado), zona geográfica en la que reside, y nivel de educación. Los factores demográficos considerados son: edad del pensionista, y género.

La diferencia en los costos de vivienda entre Lima y otras regiones del país puede tener un impacto significativo en la cantidad real de dinero que un pensionista puede gastar; en Lima, los costos pueden ser más altos, y si un pensionista de la AFP (con una pensión promedio más alta) vive en Lima, sus gastos de vivienda podrían consumir una mayor proporción de su pensión que alguien en la ONP viviendo fuera de Lima, esto crea una disparidad en la cantidad de dinero disponible para otros gastos esenciales.

Los gastos del hogar, como servicios públicos y mantenimiento, también varían según la región y el tipo de propiedad en la que vive el pensionista; la capacidad de mantener una calidad de vida aceptable dependerá de

cuánto de su pensión se destine a estos gastos, la inequidad puede surgir de diferencias en la eficiencia y costo de los servicios en diferentes regiones, así como de diferencias entre los dos subsistemas de pensiones, con los pensionistas de la AFP posiblemente teniendo más capacidad para cubrir estos costos.

Los costos de atención médica y la calidad de los servicios de salud también pueden variar significativamente en Perú, y esto se refleja en la equidad entre los pensionistas; los pensionistas de la AFP podrían tener acceso a mejores servicios de salud privados o mayores ahorros para cubrir estos costos, mientras que los de la ONP podrían depender más de los servicios de salud públicos, que pueden variar en calidad y disponibilidad según la región.

Estos diferentes factores contribuyen a la inequidad entre los pensionistas en Perú al crear variaciones en los gastos esenciales en función de la ubicación y del subsistema de pensiones al cual pertenecen; la vivienda, el hogar y la salud son áreas vitales que consumen una proporción significativa de los ingresos de una persona jubilada, y las diferencias en estos gastos se magnifican por las disparidades regionales y las diferencias entre los sistemas de AFP y ONP, la combinación de estos factores puede dar lugar a una experiencia muy diferente de la jubilación según dónde y cómo vive una persona, y qué tipo de pensión recibe. Esto podría ser una llamada a revisar cómo el sistema de pensiones podría ser reformado o ajustado para reducir estas disparidades y asegurar una jubilación más equitativa para todos.

CONCLUSIONES

- Se ha constatado que la equidad en la distribución de ingresos por pensiones en el sistema previsional peruano (2012-2022) ha sido significativamente influenciada por factores socioeconómicos y demográficos específicos, lo que ha resultado en disparidades notables. Este hallazgo subraya la complejidad del sistema previsional y la necesidad de considerar múltiples dimensiones en la formulación de políticas.
- Los pensionistas en diferentes zonas geográficas experimentan variaciones significativas en sus ingresos por pensiones. Esta disparidad regional indica la necesidad de políticas que aborden específicamente las desigualdades geográficas para garantizar una distribución más equitativa de los recursos previsionales.
- La naturaleza de la entidad laboral (pública o privada) donde trabajó el pensionista influye considerablemente en la cuantía de sus pensiones. Este resultado sugiere que las diferencias estructurales entre los sectores público y privado tienen un papel clave en la determinación de los beneficios previsionales.
- Se observa una correlación significativa entre el tipo de sistema previsional (público o privado) receptor de los aportes y la cantidad de ingresos que recibe el pensionista. Esto resalta la importancia de evaluar y potencialmente reformar las estructuras y políticas de ambos sistemas para lograr una distribución más equitativa.
- El género del pensionista es un determinante clave en la disparidad de los ingresos por pensiones. Este hallazgo evidencia la existencia de brechas de género en el sistema previsional y la necesidad de implementar medidas para abordar estas desigualdades.
- El nivel educativo alcanzado por los pensionistas afecta significativamente sus ingresos por pensiones. Este aspecto subraya la interconexión entre la educación y la seguridad económica en la vejez, sugiriendo la importancia de políticas educativas y laborales inclusivas.

REFERENCIAS BIBLIOGRÁFICAS

- Alonso, J., Sánchez, R., & Tuesta, D. (2014). Un modelo para el sistema de pensiones en el Perú: Diagnóstico y recomendaciones. *Estudios Económicos*(27), 81-98.
- Arrieta, A., Jaramillo, M., Prieto, L., Seinfeld, J., De la Torre, A., & Tuesta, D. (2017). *Propuestas de reformas en el sistema de pensiones, financiamiento en la salud y seguro de desempleo*. Lima: Comisión de Protección Social.
- Asociación de AFP. (13 de Setiembre de 2022). *Asociación de AFP: Estadísticas*. <http://estadisticas.asociacionafp.pe/ActiveAffiliatesAccordingAfp>
- Banco Mundial. (13 de Setiembre de 2022). *Banco Mundial Datos*. <https://datos.bancomundial.org/indicador/SP.DYN.LE00.FE.IN?view=chart>
- Bernal, N. (2016). Los gastos públicos en pensiones en América Latina y sus proyecciones al año 2075: evidencia de Chile, Perú, Colombia y México. *Apuntes*, XLIII(79), 79-128. <https://doi.org/http://dx.doi.org/10.21678/apuntes.79.867>
- Breiman, L., Friedman, R., Olshen, A., & Stone, J. (1984). *Classification and Regression Trees*. Hall/CRC.
- Brower Beltramin, J. (2020). La Justicia como equidad. Una reformulación. *Lus et Praxis*, 597-600.
- Cardona, D., & Peláez, E. (2012). Envejecimiento poblacional en el siglo XXI: oportunidades, retos y preocupaciones. *Salud Uninorte*, 335-348.
- Cenas Rodríguez, S. (2019). *Desigualdad del ingreso en el Perú: 1997-2017*. Universidad Nacional de Trujillo.
- Chávez Jorge, D. (2018). *Gasto público y desigualdad de ingreso: Perú, 1997-2017*. Lima: Universidad César Vallejo.
- Collier, A. (2016). *Data Management for Researchers*.

- Congreso de la República del Perú. (5 de Mayo de 2016). LEY N° 30478. <https://busquedas.elperuano.pe/normaslegales/ley-que-modifica-el-articulo-40-y-la-vigesimo-cuarta-disposi-ley-n-30478-1398360-6/>
- Couture de Troismonts, R. (1975). *Manual de técnicas de documentación*. Marymar.
- Esping-Andersen, G. (1990). *Los tres mundos del estado del bienestar*. Edicions Alfons el Magnanim. <https://doi.org/84-7822-097-6>
- Espinoza, E. (2018). La hipótesis en la investigación. *Mendive*, 122-139. <https://doi.org/1815-7696>
- Fabrigar, L. (1999). Evaluating the use of Exploratory Factor Analysis in Psychologic Research. *Psychological Methods*, 272-299.
- Fondo de Población de las Naciones Unidas. (13 de Setiembre de 2022). *Envejecimiento de la población*. <https://www.unfpa.org/es/envejecimiento-de-la-poblaci%C3%B3n>
- Freire, P. (2005). *Pedagogía del oprimido*. Siglo XXI Editores S.A. <https://doi.org/968-23-2589-7>
- Gestión, R. (8 de Junio de 2020). Perú será el país con mayor caída de PBI en Sudamérica este año, según BM. *Gestión*, pág. 4.
- Gómez Pérez. (2015). *Introducción a la metodología de la investigación científica*.
- Hernandez Sampieri, Fernandez Collado, & Baptista Lucio. (2021). *Metodología de la Investigación*.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). *Metodología de la investigación*. McGraw-Hill. <https://doi.org/978-1-4562-2396-0>
- Hoyos Gómez, D. (2008). Elementos para una teoría de justicia: una comparación entre John Rawls y Amarthya Sen. *Desafíos*, 156-181.
- Instituto Nacional de Estadística e Informática. (2017). *Estadísticas*. Encuesta Nacional de Hogares: <https://www.inei.gob.pe/>
- Instituto Nacional de Estadística e Informática. (3 de noviembre de 2022). *ENAHO, Condiciones de vida en el Perú*. <https://m.inei.gob.pe/biblioteca-virtual/boletines/condiciones-de-vida/1/#lista>
- Instituto Nacional de Estadística e Informática. (3 de noviembre de 2022). *Ficha Técnica de la Encuesta Nacional de Hogares 2017*. <http://iinei.inei.gob.pe/iinei/srienaho/Descarga/DocumentosMetodologicos/2017-55/FichaTecnica.pdf>
- MARKA Investigación. (3 de noviembre de 2022). *Blog de MARKA Investigación*. <https://markainvestigacion.wordpress.com/>
- Marshall, T., & Bottomore, T. (2023). *Ciudadanía y clase social*. Alianza Editorial. <https://doi.org/9788411481809>
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media.
- Mesa-Lago, C. (2022). Desempeño de pensiones privatizadas en América Latina, 1980-2020: el caso de México. *El Trimestre Económico*, LXXXIX(355), 755-794. <https://doi.org/10.20430/ete.v89i355.1477>
- Ministerio de Economía y Finanzas. (2004). *Los Sistemas de Pensiones en Perú*. Lima: Dirección General de Asuntos Económicos y Sociales.
- Murphy, T. (2016). Dos teorías sobre la justicia social. *Economía*, 78-89.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Nussbaum, M. (1992). Human functioning and social justice: in defense of aristotelian essentialism. *Political theory*, 20(2), 202-246.
- Nussbaum, M. (2012). *Crear capacidades: Propuesta para el desarrollo humano*. Espasa Libros. <https://doi.org/978-84-493-0988-5>
- OCDE. (2021). *Pensions at a Glance 2021: OECD and G20 Indicators*. París: OCDE.
- Oficina de Normalización Previsional. (13 de setiembre de 2022). *Sistema previsional y sus actores*. https://www.onp.gob.pe/pensiones_peru_onp/sistema_previsional
- Organismo de Estados Americanos. (13 de setiembre de 2022). *Tratados multilaterales*. https://www.oas.org/es/sla/ddi/tratados_multilaterales_interamericanos_A-70_derechos_humanos_personas_mayores.asp
- Organización Mundial de la Salud. (9 de noviembre de 2022). *Envejecimiento y salud*. <https://www.who.int/es>
- Otero-Bahamón, S., Álvarez-Altamiranda, L., Sampayo, A., & Alvis, J. (2022). *Trayectorias divergentes de la desigualdad en las ciudades intermedias: el rol de las pensiones y las ayudas en la reducción de la desigualdad de ingresos*. Universidad del Rosario, Bogotá, Colombia. <https://doi.org/10.18046/recs.iEspecial.4932>
- Palomar, J. (9 de noviembre de 2022). *Etimologías Filosóficas*. Etimologías Filosóficas: <http://etimologiaspalomar.blogspot.com/2011/10/equidad.html>
- Peterson, R. (2000). Meta-Analysis of Variance Accounted for and Factor Loadings in Exploratory Factor Analysis. *Marketing Letters*, 261-275.
- Quinn Patton, M. (2015). *Qualitative Research & Evaluation Methods*.
- Ramirez, T. (2010). *Como hacer un proyecto de investigacion*. Caracas: Editorial Panapo. <https://doi.org/980-733903-2>
- Rawls, J. (1999). Justicia como equidad. *Revista española de control*, 129-158.
- Rawls, J. (2012). *Teoría de la justicia*. México D.F.: FCE - Fondo de Cultura Económica. <https://doi.org/607-16-0903-8>

- Rodríguez López, M. (2018). *Diseño y validación de un índice focalizado y compuesto basado en el índice de concentración para la medición de inequidad en salud*. Bogotá: Universidad Nacional de Colombia.
- Rodríguez Weber, J. (2014). *La economía política de la desigualdad de ingreso en Chile, 1850-2009*. Montevideo, Uruguay: Universidad de la República.
- Rodríguez, M. (2015). Reseña de libro *La idea de justicia* de Amartya Sen. *Recerca : revista de pensament i anàlisi*, 17(7), 129-147. <https://doi.org/10.6035>
- Rosas Febres, M. (2018). *Inequidad dentro de la equidad, pagos informales durante la atención de los afiliados al seguro integral de salud 2008 a 2010*. Universidad Peruana Cayetano Heredia.
- Sen, A. (1999). *Desarrollo y libertad*. 2000, Editorial Planeta S.A. <https://doi.org/950-49-0473-4>
- Silva, Y. (2017). *Igualdad, equidad, derecho y justicia en el régimen pensional colombiano*. Bogotá: Universidad Católica de Colombia.
- Smith, J., Johnson, A., & Rodríguez, M. (2019). Comparative analysis of programming languages for efficient data collection from structured databases. *Data Science*, 10(3), 345-362.
- Social, C. d. (2017). *Propuestas de reforma en el sistema de pensiones, financiamiento en la salud y seguro de desempleo*. Lima: CPS.
- Stilman, G. (5 de Octubre de 2022). *Justicia de justicias: Ética universal y progreso moral en un mundo plural*. https://www.academia.edu/11274307/Justicia_de_justicias_Etica_universal_y_mejoramiento_moral_en_un_mundo_plural
- Superintendencia de Banca, Seguros y AFP. (13 de Setiembre de 2022). *Comisiones y Primas de Seguro del SPP*. https://www.sbs.gob.pe/app/spp/empleadores/comisiones_spp/paginas/comision_prima.aspx
- Superintendencia de Banca, Seguros y AFP. (14 de Setiembre de 2022). *SBS para ciudadanos*. <https://www.sbs.gob.pe/usuarios/informacion-de-pensiones/otros-beneficios-del-spp/entrega-de-hasta-el-955-del-fondo-de-pensiones>
- UCLA, I. f. (16 de julio de 2020). *A practical introduction to factor analysis: Exploratory factor analysis*. <https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/>
- Unidas, N. (2020). *El impacto del COVID-19 en América Latina y el Caribe*. New York: Naciones Unidas.
- Urbano-Guzmán, M. (2014). El concepto de igualdad en algunas teorías contemporáneas de la justicia. *Criterio libre jurídico*, 123-139.
- Varea, J. (15 de Junio de 2020). ¿Cómo afectará la crisis de la Covid-19 a nuestro sistema de pensiones? *La Vanguardia*, pág. 7.
- Vickery, B. (1970). *Techniques of information retrieval*. Londres: Butterworths.
- Vidal, P. F. (2009). *La teoría de la justicia social en Rawls*. Universidad Federal de Rio de Janeiro.
- Yamada, G., Castro, J., & Oviedo, N. (2016). *Revisitando el coeficiente de Gini en el Perú: El rol de las políticas públicas en la evolución de la desigualdad*. Lima: Universidad del Pacífico.
- Yin, R. (2014). *Case Study Research, Design and Methods*.