COMILLAS
UNIVERSIDAD PONTIFICIA

Facultad de Ciencias Económicas y Empresariales
Grado en Administración y Dirección de Empresas y Business Analytics

REDES DE KOLMOGOROV–ARNOLD APLICADAS A LA ESPERANZA DE VIDA

Clave: 202011547

Autor: Álvaro Inchausti Vera
Director: Eduardo César Garrido Merchán

MADRID | Abril 2025

Agradecimientos

Me gustaría agradecer a Eduardo, mi tutor, por acompañarme con tanta disposición durante todo el proceso. Su apoyo, cercanía y guía han sido clave para sacar adelante este trabajo, y ha sido un verdadero lujo contar con alguien de su nivel en la universidad.

También quiero dar las gracias a mi familia, por su ánimo constante y por hacerme sentir acompañado en cada etapa de esta experiencia.

Índice de contenidos

1. INTRODUCCIÓN Y MOTIVACIÓN 9

2. ESTADO DEL ARTE 12

2.1. Factores socioeconómicos 13
2.2. Gasto sanitario, acceso y calidad 16
2.3. Factores medioambientales e higiénicos 18
2.4. Factores de comportamiento y estilo de vida 19
2.5. Factores relacionados con la comunidad 20

3. ALCANCE DEL TRABAJO 24

3.1. Objetivos 24
3.2. Hipótesis 25
3.3. Asunciones 25
3.4. Restricciones 26

4. METODOLOGÍA 28

4.1. Técnica 28
4.1.1. Motivación 28
4.1.2. Splines 29
4.1.3. Fundamento teórico 31
4.1.4. Arquitectura de la red KAN 33
4.1.5. Implementación práctica 37
4.2. Marco teórico de la esperanza de vida 39

5. DESCRIPCIÓN Y RESULTADOS DE LOS EXPERIMENTOS 40

5.1. Descripción 40
5.2. Implementación 43
5.2.1. Red KAN 43
5.2.2. Regresión lineal 45
5.3. Resultados 45
5.3.1. Red KAN 46
5.3.2. Regresión lineal 49
5.3.3. Comparativa 49

6. CONCLUSIONES Y RESULTADOS 52

6.1. Revisión de objetivos 52
6.2. Resumen de hallazgos 53
6.3. Implicaciones y aportaciones 54
6.4. Líneas futuras de investigación 54

7. DECLARACIÓN DE USO DE HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL GENERATIVA 56

8. BIBLIOGRAFÍA 57

9. ANEXO 60

9.1. Código completo red KAN 60
9.1. Código completo regresión lineal 67

Índice de figuras

Figura 1. Objetivos de Desarrollo Sostenible (ODS) 10

Figura 2. Determinantes principales de la salud 12

Figura 3. Relación de la esperanza de vida y el salario nacional per cápita 14

Figura 4. Complejidad de la transmisión del gasto público 17

Figura 5. Ilustración de un spline 30

Figura 6. Arquitectura de una KAN 33

Figura 7. Flujo de una KAN 34

Figura 8. Ejemplo práctico con una función conocida 35

Figura 9. Representación de la arquitectura interna de una KAN 37

Figura 10. Proceso de entrenamiento y extracción simbólica en una red KAN 38

Figura 11. Arquitectura KAN 47

Figura 12. Modelo KAN podado 48

Figura 13. Predicciones regresión lineal 50

Figura 14. Predicciones KAN 50

Índice de tablas

Tabla 1. Resumen del estado del arte 22

Tabla 2. Resumen de las variables incluidas en la base de datos 42

Índice de ecuaciones

Ecuación 1. Representación de un spline 29

Ecuación 2. Combinación lineal de funciones base 31

Ecuación 3. Representación funcional general de una red KAN..... 31

Resumen

Este trabajo tiene como objetivo principal modelizar la esperanza de vida en función de determinadas variables socioeconómicas y demográficas, aplicando para ello una red Kolmogorov–Arnold (KAN), una arquitectura reciente que permite obtener expresiones simbólicas interpretables. Como referencia, se ha utilizado un modelo de regresión lineal con efecto cuadrático, lo que permite comparar ambos enfoques desde el punto de vista del ajuste y la interpretabilidad.

Para alcanzar este objetivo, se ha llevado a cabo una revisión de la literatura científica con el fin de identificar los principales determinantes de la esperanza de vida. A partir de esta revisión, se ha construido una base de datos que ha permitido entrenar los modelos. El modelo KAN se ha ajustado mediante optimización bayesiana.

Los resultados muestran que la red KAN ofrece un rendimiento superior al modelo lineal, alcanzando un R^2 de 0,846 frente a 0,699 y un MSE de 5,55 frente a 10,84. Sin embargo, este mayor rendimiento conlleva una pérdida de simplicidad en la expresión resultante, lo que plantea un interesante debate entre precisión e interpretabilidad.

Además, el estudio sugiere que el uso de modelos más transparentes, como las redes KAN, puede contribuir al desarrollo de políticas públicas más equitativas y alineadas con los Objetivos de Desarrollo Sostenible (ODS), en particular en lo relativo a salud, igualdad y gobernanza. En definitiva, el trabajo evidencia el potencial de las redes KAN para abordar problemas de modelado simbólico en contextos sociales, abriendo la puerta a futuras líneas de investigación.

Palabras clave: esperanza de vida, red Kolmogorov–Arnold, KAN, modelado simbólico, optimización bayesiana, interpretabilidad de modelos, justicia.

Abstract

This study aims to model life expectancy based on specific socioeconomic and demographic variables by applying a Kolmogorov–Arnold Network (KAN), a recent architecture that enables the extraction of interpretable symbolic expressions. As a benchmark, a quadratic-effect linear regression model has been used, allowing for a comparison between both approaches in terms of predictive performance and interpretability.

To achieve this objective, a review of the scientific literature was conducted to identify the main determinants of life expectancy. Based on this review, a database was built to train the models. The KAN model was tuned using Bayesian optimization.

The results show that the KAN outperforms the linear model, achieving an R^2 of 0.846 compared to 0.699, and an MSE of 5.55 versus 10.84. However, this improved performance comes at the cost of a more complex symbolic expression, raising an interesting discussion about the trade-off between accuracy and interpretability.

Furthermore, the study suggests that the use of more transparent models, such as KANs, can contribute to the development of fairer public policies aligned with the Sustainable Development Goals (SDGs), particularly those related to health, equality, and governance. Ultimately, this work highlights the potential of KANs to address symbolic modeling challenges in social contexts, opening the door to future research directions.

Keywords: life expectancy, Kolmogorov–Arnold Network, KAN, symbolic modeling, Bayesian optimization, model interpretability, justice.

1. Introducción y motivación

La esperanza de vida constituye un indicador esencial del desarrollo de una sociedad, al sintetizar los efectos conjuntos del progreso tecnológico y social. Según la Organización para la Cooperación y el Desarrollo Económicos (2023), “la esperanza de vida al nacer mide cuánto tiempo vivirían, en promedio, las personas, en función de un conjunto determinado de tasas de mortalidad específicas por edad” (p. 62) [traducción propia]. La persistente disparidad de este indicador entre países suscita interrogantes sobre los factores subyacentes, los cuales han sido objeto de estudio y han motivado la literatura en el último medio siglo.

A pesar de los importantes avances sociales y económicos, múltiples estudios constatan que las desigualdades en esperanza de vida persisten incluso en los estados de bienestar más consolidados. Mackenbach (2012) analiza en profundidad esta paradoja, señalando que, aunque las políticas redistributivas han mejorado las condiciones materiales, las desigualdades en salud no han desaparecido e, incluso, han tendido a ampliarse en las últimas décadas. Estas desigualdades no solo afectan a la esperanza de vida, sino también a los años de vida saludable, y tienen su origen en un complejo entramado de factores que incluyen desde la educación y el nivel socioeconómico hasta el acceso a recursos culturales.

Sin embargo, la mayoría de estudios hasta la fecha que abordan este fenómeno se caracterizan por tener un enfoque principalmente cualitativo o por modelizar la esperanza de vida con variables aisladas. En el primer caso, los estudios ofrecen marcos conceptuales sólidos que identifican los determinantes clave, pero carecen de una formalización matemática que permita estimaciones precisas. En el segundo caso, los análisis cuantitativos tienden a centrarse en relaciones de dos variables o en modelos lineales con capacidad explicativa limitada. Este enfoque dificulta la comprensión de los efectos simultáneos e interrelacionados que múltiples variables ejercen sobre la esperanza de vida.

En este contexto, las redes Kolmogorov-Arnold (KAN) surgen como una alternativa prometedora. Esta arquitectura ofrece la posibilidad de construir modelos simbólicos que no solo alcanzan un buen rendimiento explicativo, sino

que también mantienen la interpretabilidad funcional de las relaciones entre variables. A diferencia de otros modelos de aprendizaje automático de tipo “caja negra”, las KAN permiten generar expresiones matemáticas explícitas que pueden ser fácilmente analizadas y utilizadas como base para la formulación de políticas públicas.

Figura 1. Objetivos de Desarrollo Sostenible (ODS), reproducido de Naciones Unidas

Esta característica resulta especialmente relevante si se considera la agenda global actual en torno a los Objetivos de Desarrollo Sostenible (ODS), en particular el ODS 3 (Salud y bienestar), el ODS 10 (Reducción de las desigualdades) y el ODS 16 (Instituciones sólidas y transparentes). Comprender de forma más profunda y transparente los determinantes estructurales de la esperanza de vida no solo mejora la calidad del análisis académico, sino que también puede contribuir a la elaboración de políticas más justas, informadas y efectivas.

Cutler, Deaton y Lleras-Muney (2006) sostienen que las mejoras históricas en la esperanza de vida han sido el resultado de la interacción de múltiples factores: el crecimiento económico, las mejoras en salud pública, la expansión de la medicina y cambios en los comportamientos individuales. Además, las desigualdades dentro de un país, especialmente las relacionadas con renta y educación, siguen siendo determinantes cruciales de las diferencias en salud. Su trabajo sugiere que la esperanza de vida es el resultado de procesos acumulativos y no lineales donde la interacción entre ingresos, entorno sanitario, comportamiento individual y contexto social no puede ser ignorada.

A la luz de esta evidencia, surge la necesidad de explorar herramientas analíticas capaces de representar adecuadamente esta complejidad multivariable y no lineal. En este sentido, el presente estudio tiene como objetivo arrojar luz sobre este vacío en la literatura aplicando redes KAN, con el fin de modelizar la esperanza de vida mediante una expresión simbólica que permita interpretar de forma explícita la interacción entre las variables que la determinan. Posteriormente, se comparan los hallazgos obtenidos con el estado actual del arte, con el propósito de evaluar la coherencia de este nuevo enfoque respecto a estudios previos y así contribuir al debate académico.

Más allá de lo innovador que pueda llegar a resultar la aplicación de las redes KAN en este campo, existen otras fuentes de motivación que alimentan este texto. La esperanza de vida es un tema que afecta a todas las personas y comprender cuáles son las causas subyacentes de una manera más profunda permite mejorar las condiciones de vida de los individuos. Además, un entendimiento más extenso mejora la toma de decisiones a nivel individual, para configurar un estilo de vida saludable, e institucional, para diseñar estrategias que promuevan la salud pública y la calidad de vida de la población.

Asimismo, diferentes estudios desmienten las repetidas aclamaciones históricas que han realizado los expertos en relación al supuesto techo al que ha llegado la esperanza de vida. Durante los últimos 160 años, ha aumentado de manera estable el número de años promedio que viven las personas (Oeppen y Vaupel, 2002). Por tanto, existe una oportunidad para seguir avanzando hacia una sociedad más saludable y longeva.

La estructura del documento se organiza de la siguiente manera: en primer lugar, se presenta una revisión del estado del arte que justifica la selección de variables y proporciona el marco contextual del estudio. A continuación, se define el alcance del trabajo especificando los objetivos, hipótesis y limitaciones. Seguidamente, se describe la metodología empleada y se exponen los resultados obtenidos. Finalmente, se presentan las conclusiones y se proponen posibles líneas futuras de investigación.

2. Estado del arte

La siguiente sección tiene por objetivo recoger las aportaciones académicas más relevantes y relacionadas con esta investigación que se han realizado alrededor de la esperanza de vida. En primer lugar, se presentará un estudio que aporta una visión holística del fenómeno y establece un marco conceptual que servirá como referencia para establecer una clasificación temática de la literatura. Al final de la sección se incluye la tabla 1 que ordena cronológicamente los estudios recogidos.

Uno de los marcos de trabajo más citados e influyentes en el campo de la salud y esperanza de vida es el propuesto por Dahlgren y Whitehead (1991). Para abordar las políticas sanitarias, los autores proponen un modelo que describe los factores que afectan de manera positiva o negativa en la salud de las personas. Lo que resulta realmente ilustrador de su aportación es la categorización de estos factores en capas de influencia consecutivas que permiten entender de manera concisa y visual los distintos determinantes.

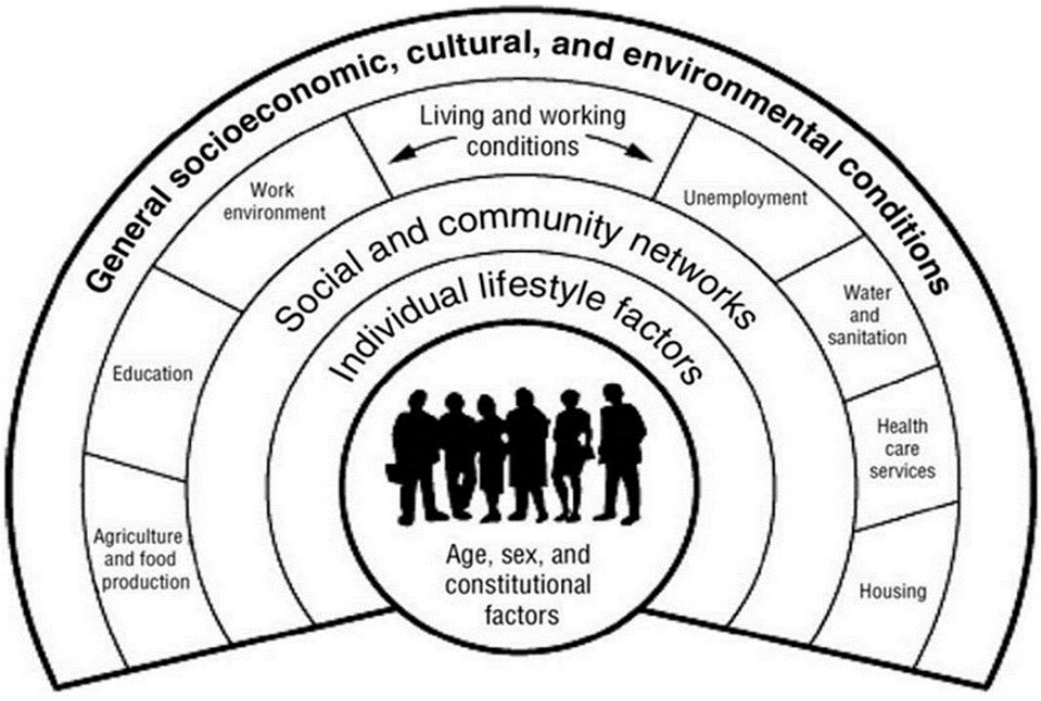

Figura 2. Determinantes principales de la salud, reproducido de *Policies and strategies to promote social equity in health* por G. Dahlgren y M. Whitehead, 1991

El modelo clasifica los factores en cuatro niveles: en la capa externa, se localizan los factores estructurales socioeconómicos y medioambientales. La siguiente capa refleja las condiciones en las que viven y trabajan las personas, que a su vez esta determinadas por otras variables como el desempleo, el alojamiento o la educación. A continuación, se encuentran factores relacionados con el apoyo recibido por la comunidad que rodea a los individuos. Por último, se muestran los determinantes individuales que se manifiestan en hábitos como el consumo de alcohol y tabaco. Fuera de estos niveles quedan factores como la edad o el género, que aun influyendo en la salud de las personas y en consecuencia, a la esperanza de vida, son fijos y no se pueden cambiar.

Esta clasificación proporciona una introducción al estado del arte, ya que ofrece una visión general del fenómeno y ha sido ampliamente referenciada en investigaciones posteriores. En 2009, Bambra et al. realizaron una revisión sistemática sobre las intervenciones destinadas a reducir las desigualdades en salud, utilizando este mismo marco de trabajo, como han hecho otros investigadores de este ámbito. Esto demuestra su relevancia y utilidad como herramienta analítica en el estudio de los determinantes de la salud.

Las siguientes subsecciones tienen por objetivo discutir diferentes aportaciones al estudio de la esperanza de vida tomando como referencia el modelo de Dahlgren y Whitehead. Esta clasificación no busca ser exhaustiva, sino más bien pragmática, para desarrollar conceptos clave en la selección de variables del modelo KAN.

2.1. Factores socioeconómicos

Esta categoría es sin duda la más estudiada y probablemente arroje los indicadores que explican con mayor fuerza la esperanza de vida. Se puede estudiar el impacto de factores socioeconómicos desde múltiples perspectivas. A continuación se expone como distintos autores han relacionado la salud de las personas con el nivel de ingresos, el grado de desigualdad social y la preparación académica.

Según Preston (1975), el salario nacional per cápita es el mejor indicador de la calidad de vida de las personas. Además, esta cifra es especialmente relevante en el estudio de determinantes de la salud dado que es ampliamente utilizada a la hora de determinar políticas sanitarias. En su estudio, analiza utilizando cohortes la relación de la esperanza de vida en el nacimiento con el ingreso nacional per cápita ajustado. Preston concluye que las curvas que reflejan la relación se han desplazado verticalmente, lo que demuestra un aumento generalizado de la esperanza de vida entre distintos grupos de años, y que la mortalidad no se ha disociado de los ingresos a lo largo del tiempo.

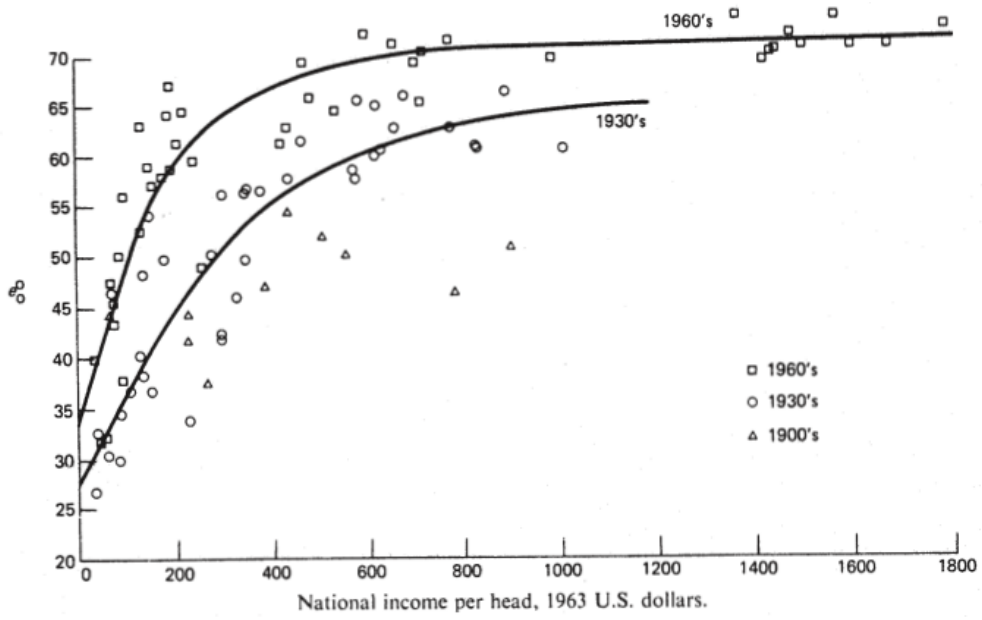

Figura 3. Relación de la esperanza de vida y el salario nacional per cápita, reproducido de *The changing relation between mortality and level of economic development* por S. Preston, 1975

Preston explica además como otros autores han tratado de relacionar la esperanza de vida con la mortalidad. En primer lugar, se puede establecer un vínculo entre el valor absoluto de la mortalidad y los ingresos en un momento determinado del tiempo. En segundo lugar, el nivel de ingresos se puede comparar con la tasa de variación de la mortalidad. Por último, hay autores que utilizan la variación de ambas variables simultáneamente. Independientemente de la modalidad utilizada, los estudios pretenden demostrar una relación indirecta entre variables, para el primer ejemplo, un mayor nivel de ingresos

resultaría en un menor número de mortalidad. Sin embargo, estas relaciones no tienen una peso estadístico tan fuerte como la elegida por Preston.

Otra escuela de investigadores ha centrado sus esfuerzos en explicar el impacto que tiene la desigualdad social, medida principalmente en términos económicos. Uno de los primeros trabajos que explica desde un enfoque empírico esta relación es el de Ben-Shlomo, White y Marmot (1996). En este estudio se busca determinar el grado de relación entre mortalidad y desigualdad socioeconómica. Los resultados confirman la hipótesis inicial, la falta de recursos es un determinante del nivel de mortalidad de una población.

En 1997, Wilkinson profundiza en esta dirección y si bien reconoce este vínculo entre desigualdad y mortalidad, realiza una nueva distinción. Existe una fuerte relación entre desigualdad económica y salud dentro de una sociedad que no se mantienen al comparar poblaciones. Dicho de otra forma, las desigualdades relativas dentro un sociedad explican las diferencias en salud pero el grado de desigualdad entre poblaciones no tiene poder explicativo para determinar las diferencias de salud. Este hallazgo tiene dos implicaciones: la primera, la desigualdad económica no es un buen indicador de las diferencias en esperanza de vida entre países; la segunda, dentro de una población, el salario relativo es más relevante que el absoluto para explicar las diferencias en longevidad.

Este razonamiento es especialmente relevante, ya que plantea interrogantes sobre las causas subyacentes de cómo las desigualdades económicas dentro de una población impactan en la salud. Wilkinson, en su trabajo, explica que las personas en la cola de la izquierda de ingresos suelen sufrir un nivel de estrés más elevado y viven en vecindarios menos favorecidos con menos cohesión y más exposición a la violencia.

La aportación de Marmot (2005) permite entender en mayor profundidad las ideas de Wilkinson. Marmot apunta que más allá del estrés o la exclusión social que pueden sufrir los individuos más desfavorecidos existen otros factores que afectan a la salud. Los más destacables son el desempleo, la carencia de apoyo social y las adicciones, condiciones más frecuentes en este grupo. Todos estos motivos apoyan la afirmación que relaciona la desigualdad dentro de una sociedad con una divergencia en esperanza de vida.

Para concluir esta sección, se recoge también el trabajo de investigadores que han abordado este fenómeno tomando como punto de partida la educación. Según Cutler (2006), la relación entre educación y salud es bidireccional, ambas son a la vez causa y efecto de la otra. Un nivel de salud más bajo se relaciona con una peor educación y al mismo tiempo, un aumento en la educación mejoraría la salud de las personas. Por tanto, un individuo podría entrar en un círculo vicioso, una mala educación generaría una peor salud que a su vez generaría una educación aún más deficiente.

Cutler apunta que la educación es un concepto que se relaciona a su vez con el nivel de ingresos, un acceso a mejores trabajos y un aumento de pensamiento crítico. Es decir, tener una mejor educación va a estimular el pensamiento crítico del individuo quien teóricamente tendría acceso a mejores trabajos que estarían mejor remunerados. Un mayor salario mejoraría el acceso de los individuos a una sanidad de calidad que justificaría una mayor esperanza de vida promedio. Esta aportación a literatura hila perfectamente con todos los razonamientos que se han expuesto en esta sección. El impacto socioeconómico en la esperanza de vida está ampliamente demostrado, independientemente de la metodología.

2.2. Gasto sanitario, acceso y calidad

Parece trivial relacionar el gasto público sanitario o el grado de acceso y calidad del sistema sanitario de un país con la salud de sus habitantes. El acceso a la sanidad se ha reivindicado históricamente como un derecho esencial del ser humano. Un ejemplo de estos manifiestos es la declaración Alma-Ata de 1978 por la Organización Mundial de la Salud, resultado de una conferencia sobre atención médica primaria. Los participantes solicitan la colaboración de los distintos sectores, no solo del ámbito sanitario, para garantizar unas condiciones de vida saludables. Esto no debería sorprender, ya que, como se adelantaba al inicio de este capítulo, son muchos los factores que influyen en la salud de las personas.

La literatura arroja resultados mixtos en lo que se refiere a la influencia del gasto público sanitario. El estudio de Filmer y Pritchett de 1999 permite entender la complejidad de este fenómeno. Cabe resaltar que los autores analizan además

del gasto público, otras variables que son relevantes para determinar la mortalidad. Los resultados obtenidos para nivel de salario, grado de desigualdad y desarrollo educativo es consistente con lo que se expuso en la sección anterior, todas estas variables disminuyen sin lugar a dudas la mortalidad. Sin embargo, no es tan fácil medir la influencia del gasto sanitario.

El impacto empírico del gasto público en la mortalidad tiene una relación estadística muy débil, siendo el coeficiente asociado a esta variable no significativo. Los resultado de Firlmer y Pritchett siguen la misma tendencia que trabajos anteriores relacionados. Son muy pocos los estudios que realmente muestran una relación fuerte entre estos conceptos sin forzar los datos. Las conclusiones estadísticas son completamente contraintuitivas, el gasto sanitario parece estar relacionado de manera inequívoca con la salud de la población.

Ante tal controversia, los autores explican la complejidad adicional que conlleva el gasto público. Una mayor partida del presupuesto nacional de un país no se traduce en una mejora inmediata de la salud de las personas porque el dinero invertido tiene que pasar por numerosos filtros antes de tener un impacto real. Primero, hay que tener en cuenta la eficacia del sector público en su conjunto. En segundo lugar, hay que medir el grado en el que el sector público es relevante, especialmente en comparación con el sistema privado. A continuación, es necesario determinar cuál es la demanda del tratamiento que se puede ofrecer gracias al incremento de presupuesto, así como la oferta actual en el sistema privado. Por último, tendrá influencia la efectividad clínica particular del tratamiento. Este estudio ilustra perfectamente la cadena por la que pasa cada euro invertido en el sistema sanitario y justifica los resultados empíricos.

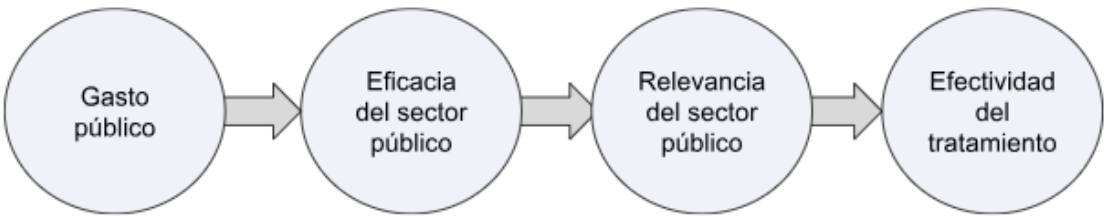

Figura 4. Complejidad de la transmisión del gasto público, elaboración propia

El trabajo de Kruk et al. (2018) permite profundizar sobre la importancia de la efectividad del sistema sanitario para garantizar una transmisión del gasto público. En este estudio se analiza el número de vidas que se podrían haber salvado en países de ingresos medios y bajos en 2016. Se estima que 8,6 millones de fallecimientos se podrían haber evitado. Del total, 3,6 millones de personas no tuvieron acceso al sistema sanitario pero 5 millones si acudieron y recibieron un cuidado deficiente que resultó en fallecimiento. Esto demuestra que independientemente de la relación estadística entre gasto público y mortalidad, un aumento de esta partida puede mejorar la salud de las personas. No obstante, es indispensable mejorar la calidad del sistema para que el impacto de la inversión pueda tener resultados eficaces.

2.3. Factores medioambientales e higiénicos

Son numerosos los determinantes que afectan a la salud relacionados con el medioambiente o higiene. Por ello, para ilustrar la importancia de esta familia de factores se describirán dos fenómenos que han sido ampliamente estudiados. El primero se refiere a la contaminación del aire, medida a través de la concentración de partículas en suspensión. El segundo abarca los efectos del consumo de agua contaminada y la falta de higiene, factores que pueden provocar enfermedades como la diarrea.

La Organización Mundial de la Salud advierte que la contaminación del aire es un factor de riesgo que causa la muerte de millones de personas cada año. En 2019, esta cifra llegó a 4,2 millones. Por lo general, el grado de contaminación se mide a través de la concentración de material particulado en suspensión (PM). Estas partículas son inhaladas por las personas y en función del tamaño y la composición tienen distintos niveles de impacto. Las más finas, de un tamaño aproximado de 2,5 micrómetros, pueden penetrar al torrente sanguíneo causando enfermedades cardiovasculares. El origen de las partículas finas se relaciona principalmente con la combustión y las reacciones químicas entre gases.

La evidencia en torno al impacto negativo de las partículas finas se recoge en estudios como el de Pope et al. de 2002. Este grupo de investigadores es capaz

de demostrar la asociación entre las partículas de 2,5 micrómetros y un riesgo más elevado de padecer enfermedades cardiopulmonares y cáncer de pulmón. Las zonas con mayor concentración de estas partículas se relacionan con una mayor mortalidad. Este es un ejemplo claro de un determinante que afecta a la esperanza de vida y se deriva directamente de las condiciones medioambientales del entorno del individuo.

Otras líneas de investigación han tratado de relacionar la contaminación en el aire con poblaciones más desfavorecidas. Sin embargo, no se puede vincular de manera directa el nivel socioeconómico con el grado de concentración de partículas puesto que las poblaciones más deprimidas no siempre viven siempre en zonas de mayor exposición (Deguen y Zmirou-Navier, 2010). No obstante, se puede argumentar que la contaminación aérea afecta en mayor medida a los grupos más vulnerables. Como se ha visto anteriormente, el acceso sanitario puede ser menor en estos colectivos y por tanto, la accesibilidad para tratar las consecuencias derivadas de la exposición sea menor.

Otro factor ambiental clave que influye en la salud es la calidad del agua y las condiciones de higiene, ya que el acceso a agua potable y a sistemas de saneamiento adecuados resulta fundamental para prevenir enfermedades infecciosas. En este sentido, Prüss-Ustün (2014) señala que el consumo de agua en mal estado, la falta de saneamiento y una higiene deficiente incrementan significativamente el riesgo de mortalidad. De hecho, solo en 2012, se estima que mejorar el acceso a agua potable y saneamiento en países de ingresos medios y bajos podría haber evitado más de 700.000 muertes. Así, tanto la contaminación del aire como la del agua se consolidan como factores de riesgo que elevan la mortalidad y, en consecuencia, reducen la esperanza de vida.

2.4. Factores de comportamiento y estilo de vida

Al profundizar en los determinantes individuales que afectan a la esperanza de vida, la última capa del marco inicial, se puede comprender mejor cómo las decisiones individuales afectan a la salud de las personas. Según Tominaga (1999), existen numerosos factores de riesgo evitables que aumentan la probabilidad de desarrollar distintos tipos de cáncer y, en consecuencia, elevan

el riesgo de mortalidad. Entre ellos se encuentran, como se adelantaba al principio del capítulo, el consumo de alcohol y tabaco. Sin embargo, hay otros factores que también influyen, como la exposición excesiva a la luz solar, a sustancias químicas o a radiación en el entorno laboral; una alimentación poco saludable, rica en grasas saturadas, calorías y sal, y deficiente en fibra, frutas y verduras; así como el estrés, cada vez más común en las sociedades occidentales.

Mención especial merece la obesidad, definida por muchos investigadores como la “pandemia del siglo XXI”. Esta enfermedad es resultado de otros factores como un estilo de vida sedentario o una nutrición no adecuada, y además, por sí misma, constituye un factor de riesgo propio. La obesidad se mide comúnmente mediante el índice de masa corporal (IMC), el cual permite clasificarla por grados, lo que facilita su análisis y la aplicación de tratamientos adecuados para cada persona. Son el segundo y tercer grado de obesidad los que empíricamente se relacionan con una mayor mortalidad (Flegal, Kit, Orpana y Graubard, 2013). Esto se explica con un riesgo más elevado de padecer diabetes o enfermedades cardiovasculares en las personas que sufren de sobrepeso.

En 2013, Ezzati y Riboli hacen una reflexión en torno a un concepto denominado como “transición epidemiológica”. Este se refiere a una tendencia generalizada entre sociedades en las que una progresiva reducción de la mortalidad relacionada con las enfermedades infecciosas, resulta en un aumento de las enfermedades no transmitibles como las cardiovasculares o el cáncer. En los últimos años, los países de ingresos medios y bajos, tradicionalmente afectados principalmente por enfermedades infecciosas, han comenzado a padecer también enfermedades no transmisibles. Esto implica que las conductas de riesgo como el consumo de alcohol y tabaco ya no solo están en los países de mayores ingresos sino también en aquellos en desarrollo.

2.5. Factores relacionados con la comunidad

Como se ha visto en este capítulo, la desigualdad económica dentro de una sociedad sirve como predictor de la esperanza de vida. Esto se explica porque una brecha económica deriva en segregación residencial que tiene implicaciones

como un aumento de la delincuencia o estrés. Esta sección trata de profundizar en el racional de estas afirmaciones a través del concepto de cohesión social

Según Kawachi y Kennedy (1997), un aumento en la solidaridad y la cohesión social mejora la salud de las personas, pero una disminución en estos factores también puede deteriorarla. Las personas que tienen más interacciones se les relaciona con un incremento de la esperanza de vida. Sin embargo, aquellos individuos que viven aislados tienen tasas de mortalidad de 2 a 3 veces más elevadas que aquellos con mayor socialización. El apoyo comunitario se presenta como otro determinante a tener en cuenta que afecta a la salud de las personas.

La cohesión social no se presenta como un indicador directo de salud, pero su influencia es innegable. Para poder analizarla adecuadamente, es necesario contar con métodos de medición específicos. En este sentido, el libro *Making Democracy Work* de Putnam, Leonardi y Nonetti (1994) ofrece una aproximación innovadora al evaluar el “capital social” de distintas regiones italianas, entendido como la densidad de participación ciudadana en actividades comunitarias, tales como asociaciones deportivas o culturales.

Los autores concluyen que las regiones con mayor capital social tienden a contar con gobiernos más funcionales, caracterizados por mayores niveles de confianza, cooperación y tolerancia ciudadana. Este tipo de gobernanza está en mejores condiciones para diseñar políticas públicas eficaces, no solo para abordar determinantes individuales, sino también para mejorar el acceso, la calidad y la eficiencia de los servicios sanitarios.

Este caso ilustra cómo, aunque los determinantes de la esperanza de vida puedan clasificarse en categorías, en la práctica están profundamente interconectados. Comprender su impacto exige adoptar una visión sistémica y multidimensional que refleje la complejidad real del fenómeno.

Tabla 1

Resumen del estado del arte

Año	Cita	Principales resultados
1975	Preston	El ingreso nacional per cápita es el mejor indicador para explicar la esperanza de vida en el momento del nacimiento.
1994	Putnam, Leonardi y Nonetti	Las regiones con mayor cohesión entre ciudadanos derivan en gobiernos más funcionales.
1996	Ben-Shlomo, White y Marmot	La mortalidad se relaciona directamente con la privación o falta de recursos.
1997	Wilkinson	Hay una fuerte relación entre la desigualdad de ingresos y la salud dentro de una población, pero no entre distintas poblaciones.
1997	Kawachi y Kennedy	Un aumento de solidaridad y cohesión social resulta en mejores condiciones de salud.
1999	Filmer y Pritchett	Vincular un mayor gasto público con una mejora en las condiciones de salud es complejo porque intervienen muchos factores adicionales.
1999	Tominaga	Los factores individuales tienen una influencia directa en el riesgo de padecer cáncer y aumentar la mortalidad.
2002	Pope et al.	La concentración de partículas finas o PM2,5 se relaciona con un mayor riesgo de mortalidad.
2005	Marmot	Profundiza sobre las causas subyacentes de la desigualdad social.
2006	Cutler	El nivel educativo influye en las oportunidades laborales, el salario y, por ende, la salud.

2010	Deguen y Zmirou-Navier	Las personas en situación de mayor vulnerabilidad son las más impactadas por la contaminación del aire.
2013	Flegal, Kit, Orpana y Graubard	La obesidad medida en función del índice de masa corporal se relaciona con un mayor riesgo de mortalidad.
2013	Ezzati y Riboli	Los factores de riesgo para enfermedades no transmisibles están creciendo en países de bajos y medianos ingresos.
2014	Prüss-Ustün	El consumo de agua contaminada y la falta de saneamiento representan un riesgo significativo para la salud.
2018	Kruk et al	El acceso a un sistema sanitario no es suficiente, es necesario garantizar unas condiciones de calidad determinadas para su efectividad.

Elaboración propia a partir de literatura científica revisada

3. Alcance del trabajo

En este capítulo se concreta el alcance de la investigación delimitando los objetivos, hipótesis, asunciones y restricciones que acompañan al trabajo. Estos elementos serán clave para enmarcar tanto la metodología empleada como la posterior interpretación de los resultados obtenidos mediante la aplicación de las redes KAN y la comparativa con los modelos tradicionales.

3.1. Objetivos

El principal objetivo de este trabajo es modelizar la esperanza de vida utilizando redes Kolmogorov-Arnold y evaluar su rendimiento frente a modelos clásicos, como la regresión lineal. Para alcanzar esta meta general, se han definido los siguientes objetivos específicos, que servirán de guía a lo largo de la investigación:

1. Revisar la literatura científica relevante relacionada con la esperanza de vida y sus determinantes, con el fin de establecer un marco conceptual sólido y justificar la selección de variables explicativas.
2. Construir un conjunto de datos adecuado que integre las variables identificadas previamente, asegurando su calidad y representatividad para el entrenamiento de los modelos.
3. Implementar y entrenar una red KAN capaz de modelizar la esperanza de vida a partir de las variables seleccionadas, explorando las capacidades simbólicas y predictivas de esta arquitectura.
4. Desarrollar un modelo de referencia basado en regresión lineal que permita comparar el rendimiento y la capacidad explicativa frente al modelo KAN.
5. Evaluar y comparar ambos modelos mediante un análisis cuantitativo, utilizando métricas como el coeficiente de determinación R^2 y el error cuadrático medio, y un análisis cualitativo centrado en la interpretabilidad de las relaciones extraídas.
6. Explorar el potencial de las redes KAN como herramienta para la generación de modelos simbólicos transparentes que puedan contribuir a

la formulación de políticas públicas más justas y alineadas con los Objetivos de Desarrollo Sostenible.

3.2. Hipótesis

El presente trabajo parte de la premisa de que las redes KAN presentan un rendimiento superior al de un modelo de regresión lineal tradicional para modelizar la esperanza de vida. Para validar esta hipótesis, se plantea un contraste formal basado en dos métricas: el coeficiente de determinación R^2 y el error cuadrático medio (MSE). Ambas serán empleadas para comparar la capacidad predictiva y de ajuste de los modelos. Se definen las siguientes hipótesis:

1. Hipótesis nula (H_0): No existen diferencias significativas en el rendimiento entre la red KAN y la regresión lineal.

$$\mathrm{R}^2_{\mathrm{KAN}} \leq \mathrm{R}^2_{\mathrm{RL}} \quad \text{y} \quad \mathrm{MSE}_{\mathrm{KAN}} \geq MSE_{\mathrm{RL}}$$

2. Hipótesis alternativa (H_1): La red KAN presenta un rendimiento significativamente superior al de la regresión lineal.

$$R^2_{\mathrm{KAN}} > R^2_{\mathrm{RL}} \quad \text{y} \quad MSE_{\mathrm{KAN}} < MSE_{\mathrm{RL}}$$

El objetivo de este trabajo será comprobar si la hipótesis alternativa se cumple, es decir, si la red KAN consigue un mayor ajuste y menor error que el modelo de referencia al describir la esperanza de vida.

3.3. Asunciones

Durante el desarrollo de la investigación se han considerado varias asunciones clave que permiten garantizar la validez y coherencia de los experimentos. Estas asunciones se dividen en dos categorías: las relacionadas con las variables y las propias de los modelos.

Asunciones relacionadas con las variables:

1. Se asume que la selección de variables basada en la literatura previa es adecuada y suficiente para explicar la esperanza de vida, sin que se haya omitido ninguna variable relevante que pueda sesgar los resultados.
2. Se presupone que las fuentes de información empleadas son precisas y reflejan de forma fidedigna las condiciones socioeconómicas y de salud de los países incluidos en la base de datos.

Asunciones relacionadas con los modelos

3. Se asume que tanto la red KAN como el modelo de regresión lineal pueden aplicarse de forma equivalente sobre el mismo conjunto de datos sin que exista sesgo que favorezca a ninguno de los dos.
4. Se presupone que las redes KAN, implementadas a través de la librería de Python seleccionada, son capaces de capturar de forma efectiva relaciones complejas y no lineales presentes entre las variables explicativas y la esperanza de vida.

3.4. Restricciones

Durante el desarrollo de la investigación se han identificado varias limitaciones que han condicionado tanto el proceso de modelización como la interpretación de los resultados. Estas restricciones se agrupan en dos tipos: limitaciones de recursos y limitaciones técnicas.

Restricciones de recursos:

1. El acceso a datos socioeconómicos completos y actualizados ha sido limitado, reduciendo el número de países disponibles en la base de datos y afectando al tamaño muestral final.
2. La frecuencia de actualización de las bases de datos empleadas no garantiza que todos los países cuenten con información reciente para todas las variables seleccionadas.
3. La capacidad computacional disponible ha limitado la complejidad de las arquitecturas empleadas, condicionando el tamaño de las redes y la duración de los entrenamientos.

Restricciones técnicas:

4. Las redes KAN constituyen un enfoque reciente, cuya investigación principal fue publicada en 2024 por Liu et al., lo que implica una documentación técnica aún limitada y una comunidad de desarrollo en estado incipiente.
5. No se han identificado estudios previos que hayan aplicado redes KAN específicamente al modelado de la esperanza de vida, lo que dificulta la comparación directa de resultados y la validación cruzada con trabajos existentes.

4. Metodología

Este capítulo se divide en dos partes: la primera aborda la técnica utilizada para entrenar el modelo, centrada en las redes KAN; y la segunda presenta el marco teórico relacionado con la esperanza de vida. Esta división dentro del capítulo de metodología resulta necesaria, ya que la investigación combina un enfoque cuantitativo, centrado en la modelización mediante redes KAN, con el análisis de la esperanza de vida, un fenómeno complejo que puede abordarse desde múltiples perspectivas.

4.1. Técnica

A continuación se desarrollarán de una manera divulgativa los conceptos clave para entender el funcionamiento de las redes KAN. Con este objetivo, se tomará como referencia la investigación de Liu et al. (2024) para el desarrollo de esta sección, ya que se trata del estudio que ha despertado el interés por este modelo y, hasta la fecha de redacción, constituye la fuente más completa sobre el tema. Es importante señalar que una descripción exhaustiva del modelo queda fuera del alcance de este trabajo, ya que el objetivo de la investigación es distinto. Por ello, se adopta un enfoque pragmático, centrado en los aspectos necesarios para entender su aplicación.

4.1.1. Motivación

Hasta la fecha de publicación del estudio mencionado (Liu et al., 2024), las redes KAN no tenían mucha importancia en el campo del *machine learning*. Esto se debe, entre otros motivos, a que la falta de una generalización del teorema matemático impedía extender la arquitectura de las redes más allá de una profundidad de dos capas, que era la configuración original. Consecuentemente, las redes KAN eran teóricamente interesantes pero carecían de aplicación práctica porque eran muy sensibles a la maldición de la dimensionalidad. Fenómeno que ocurre cuando se trabaja con muchos datos de muchas dimensiones, es decir, con un número considerable de variables.

Los autores de la investigación logran generalizar el teorema original, lo que permite construir redes más profundas y anchas, superando así la maldición de la dimensionalidad. Además, proporcionan una librería para construir y entrenar redes KAN para ejemplos reales como es caso de la esperanza de vida. El uso de funciones de una variante y sumas permite capturar relaciones no lineales complejas entre las variables que otros modelos como la regresión lineal no detectan. Por ello, las redes KAN se plantean como una alternativa equilibrada entre interpretabilidad y precisión, lo que ha motivado la realización de este estudio.

4.1.2. Splines

Antes de explicar con más detalle la arquitectura concreta de las redes KAN y el fundamento matemático, es necesario hacer una introducción sobre un concepto relacionado, el *spline*. Se trata de una función construida por tramos, definida mediante polinomios que se unen de forma suave. Su principal objetivo es aproximar o interpolar funciones complejas de manera flexible, evitando tanto la rigidez de las rectas como las oscilaciones indeseadas de polinomios globales de alto grado (Weston, 2002).

Formalmente, un *spline* es una función que satisface:

$$S(x) = \begin{cases} p_1(x), & x \in [t_0, t_1) \\ p_2(x), & x \in [t_1, t_2) \\ \vdots & \\ p_m(x), & x \in [t_{m-1}, t_m) \end{cases}$$

Ecuación 1

donde cada $p_i(x)$ es un polinomio de grado prefijado y t_i representa los nudos que dividen el dominio. La figura 5 ilustra esta relación:

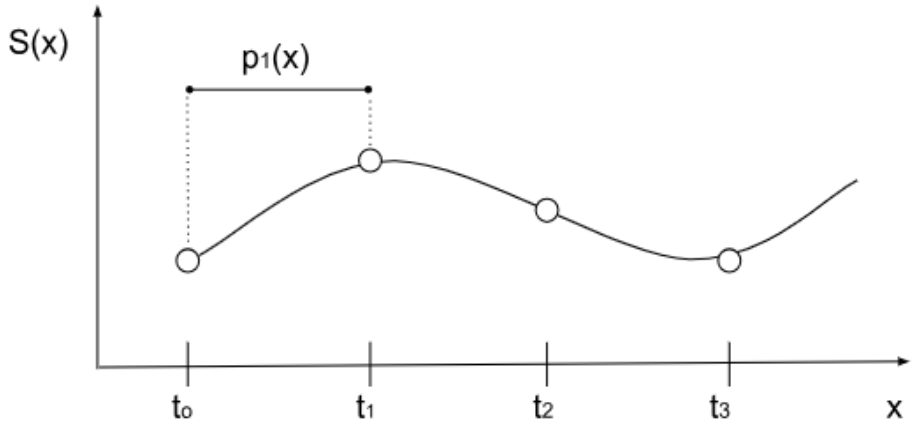

Figura 5. Ilustración de un *spline*, elaboración propia

De manera intuitiva, un *spline* puede entenderse como una “pieza” de una curva que se adapta localmente a los datos. En lugar de ajustar una única función global que describa toda la relación entre las variables, se divide el dominio en subintervalos mediante nudos. En cada subintervalo se ajusta un polinomio de grado k, asegurando que la transición entre polinomios sea suave. Este proceso da lugar a una función global continua y con derivadas continuas.

Continuidad de la función:

$$\lim_{x \to t_i^-} S(x) = \lim_{x \to t_i^+} S(x), \quad \forall i = 1, \dots, m-1$$

Continuidad de la primera derivada:

$$\lim_{x \to t_i^-} S'(x) = \lim_{x \to t_i^+} S'(x), \quad \forall i = 1, \dots, m-1$$

Continuidad de las derivadas superiores (hasta orden $r \leq k-1$)

$$\lim_{x \to t_i^-} S^{(j)}(x) = \lim_{x \to t_i^+} S^{(j)}(x), \quad \forall i = 1, \dots, m-1, \quad j = 2,3, \dots, r$$

Si el *spline* es cúbico (grado 3), entonces la continuidad se garantiza hasta la segunda derivada.

Entre las múltiples variantes de *splines*, las *B-splines* (*Basis Splines*) ocupan un lugar destacado. Su característica distintiva es que no definen directamente la función spline, sino que proporcionan una base de funciones locales sobre las que se construye la curva mediante una combinación lineal:

$$S(x) = \sum_{i=1}^{n} c_i B_i(x)$$

Ecuación 2

donde:

$B_i(x)$ son las funciones base *B-spline*

c_i son coeficientes que determinan la forma concreta de la curva

Cada una de estas funciones base solo afecta a una parte limitada de la curva, lo que otorga un control local muy valioso: modificar un coeficiente c_i solo altera la forma de la curva en la vecindad de su correspondiente *B-spline*. Esta propiedad resulta especialmente importante en redes KAN, donde las funciones que conectan las neuronas serán precisamente *B-splines* parametrizables.

4.1.3. Fundamento teórico

El modelo KAN se basa en el teorema matemático de representación de Kolmogorov-Arnold, formulado por Andrey Kolmogorov y Vladimir Arnold en el siglo XX. El teorema afirma que cualquier función continua multivariable puede representarse como la suma de funciones continuas de una sola variable.

De manera formal:

Sea $f: [0,1]^n \to \mathbb{R}$ una función continua definida sobre el hipercubo unitario. Entonces, existen funciones continuas univariantes

$\Phi_q: \mathbb{R} \to \mathbb{R}$ y $\varphi_{q,p}: [0,1] \to \mathbb{R}$ tales que:

$$f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \varphi_{q,p}(x_p) \right)$$

Ecuación 3

para todo $(x_1, x_2, \dots, x_n) \in [0,1]^n$

donde:

f es una función continua multivariable que se quiere representar

Φ_q son funciones univariantes continuas

$\varphi_{q,p}$ también son funciones univariantes continuas

Kolmogorov y Arnold probaron que para cualquier $f \in C([0,1]^n)$ es suficiente con $2n+1$ funciones exteriores Φ_q y $n(2n+1)$ funciones interiores $\varphi_{q,p}$. Este número es óptimo y no puede reducirse en general.

De forma sencilla, aunque una función pueda depender de muchas variables a la vez y parezca complicada, en realidad puede construirse sumando funciones más simples que solo actúan sobre una variable cada una. Así, en lugar de modelar directamente todas las interacciones entre las variables al mismo tiempo, es suficiente con sumar adecuadamente funciones que tratan cada variable por separado.

Sin embargo, como se señala en la investigación de Lui et al. (2024), las funciones univariantes resultantes pueden ser no suaves e incluso llegar a exhibir comportamiento fractal. Por ello, independientemente de su validez teórica, el teorema no había tenido aplicaciones prácticas en *machine learning*. Sin embargo, los autores del artículo indican que las funciones de interés suelen ser suaves y estar formadas por estructuras composicionales simples. Esta perspectiva está alineada con la filosofía propia de la física y otras ciencias aplicadas, donde el interés suele centrarse en los casos típicos y no en los peores escenarios posibles.

A modo de ilustración, a continuación se presenta un ejemplo sencillo que permite visualizar la esencia del teorema de Kolmogorov-Arnold de forma intuitiva:

Una función multivariante como $f(x_1, x_2) = x_1 x_2$, puede reescribirse, siguiendo la filosofía del teorema de Kolmogorov-Arnold, como una suma de funciones univariantes en lugar de un producto. En particular, es posible expresar:

$$f(x_1, x_2) = -\frac{x_1^2}{2} - \frac{x_2^2}{2} + \frac{(x_1+x_2)^2}{2}$$

donde:

$$f_1(x_1) = -\frac{x_1^2}{2}$$

$$f_2(x_2) = -\frac{x_2^2}{2}$$

$$f_3(x_1 + x_2) = \frac{(x_1 + x_2)^2}{2}$$

De esta forma, se representa $f(x_1, x_2)$ únicamente mediante funciones que dependen de una sola variable y su suma. El lector podrá comprobar fácilmente que, al simplificar la expresión, se recupera el producto original:

$$-\frac{x_1^2}{2} - \frac{x_2^2}{2} + \frac{(x_1 + x_2)^2}{2} = x_1 x_2$$

4.1.4. Arquitectura de la red KAN

Las redes KAN tienen una estructura similar a las redes neuronales clásicas (MLP) porque están formadas por capas completamente conectadas.

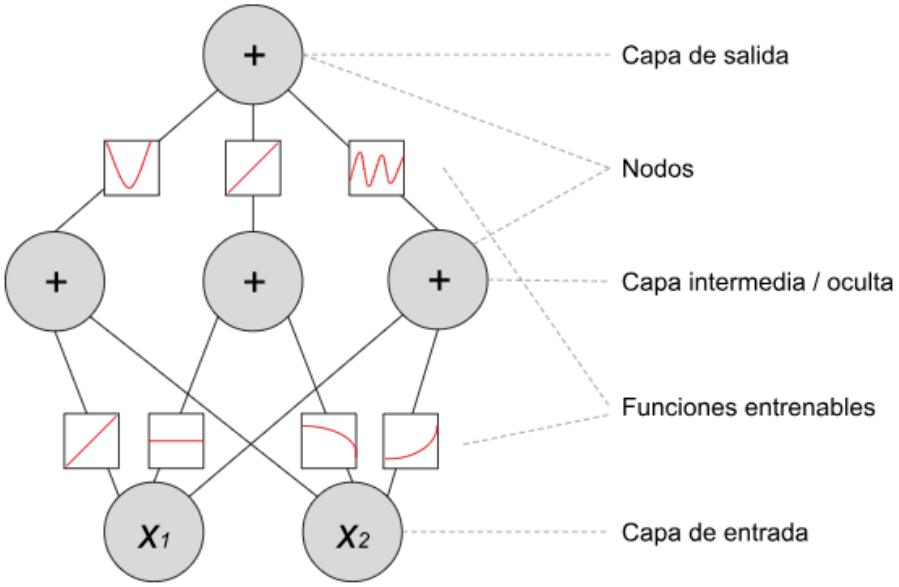

Figura 6. Arquitectura de una KAN, elaboración propia

Sin embargo, presentan diferencias esenciales con respecto a las MLP en las conexiones y los nodos. A diferencia de las redes neuronales, las conexiones entre nodos no tienen un peso escalar, sino que están asociadas a una función de una variable. Estas funciones, introducidas anteriormente en este capítulo, reciben el nombre de *splines* y son capaces de aprender dado que se ajustan durante el entrenamiento. La segunda diferencia fundamental radica en el funcionamiento de los nodos, que no aplican funciones de activación

tradicionales, sino que se limitan a sumar las salidas de las conexiones entrantes. La figura 7 ilustra de manera sencilla la arquitectura:

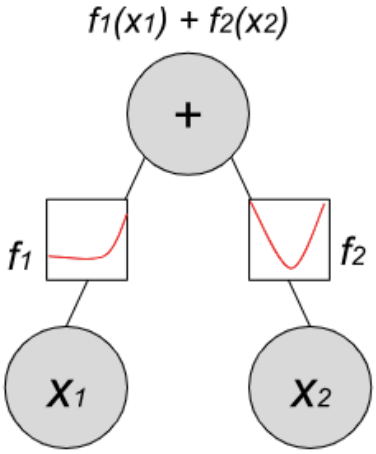

Figura 7. Flujo de una KAN, elaboración propia

El nodo, representado mediante un símbolo de suma, se encarga de combinar los resultados de las funciones univariantes. Este esquema resume de forma visual los principios básicos que definen la arquitectura de una red KAN y que serán fundamentales para el ejemplo que se presenta a continuación. Con el objetivo de reforzar la relación entre la estructura de las redes KAN y el teorema de Kolmogorov-Arnold, se muestra seguidamente la figura 8 que representa precisamente la función utilizada en el ejemplo de la sección anterior:

$$f(x_1, x_2) = x_1 x_2$$

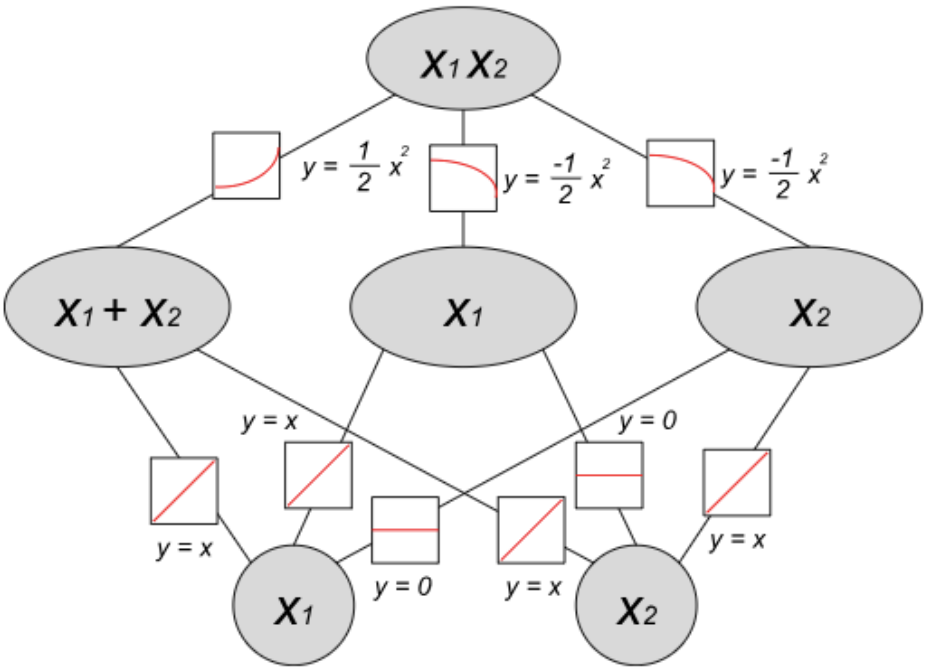

Figura 8. Ejemplo práctico con una función conocida, elaboración propia

Puede observarse cómo una red KAN implementa la función $f(x_1, x_2) = x_1 x_2$ utilizando únicamente funciones univariantes y operaciones de suma. En la parte inferior se encuentran los nodos de entrada, que representan las variables x_1 y x_2. A partir de estas entradas, se aplican funciones univariantes simples (como $y = x$ o $y = 0$) que generan combinaciones intermedias, entre ellas x_1, x_2, $x_1 + x_2$, representadas como nodos en la capa intermedia. Cada una de estas combinaciones pasa a su vez por una segunda función univariante más compleja. Finalmente, el nodo superior combina los valores resultantes y produce la salida final.

Este proceso refleja directamente la estructura del teorema de Kolmogorov-Arnold, en el que se compone una suma de funciones univariantes aplicadas a combinaciones de las variables originales. Así, la suma que se realiza en la capa intermedia corresponde con la suma interna del teorema, mientras que las funciones que actúan posteriormente sobre esas sumas representan las funciones externas Φ_q, cuya salida se agrega para obtener el valor final de la función multivariable.

La contribución principal de Liu et al. (2024) consiste en generalizar el teorema de Kolmogorov-Arnold, permitiendo su aplicación más allá de redes de dos capas. La generalización propuesta en el estudio habilita la construcción de redes profundas y anchas, en las que cada capa intermedia sigue respetando la estructura composicional del teorema. Esta ampliación teórica no solo mejora la capacidad expresiva del modelo, sino que también facilita la escalabilidad y entrenabilidad de las redes KAN en contextos de alta dimensionalidad.

Una vez comprendidos los fundamentos teóricos y la estructura básica de la arquitectura, es posible abordar cómo se entrenan las redes KAN. A diferencia de las redes neuronales tradicionales, en las que se actualizan pesos escalares que conectan los nodos, en las KAN lo que se ajustan son funciones univariantes. A primera vista, entrenar una función puede parecer poco intuitivo, pero esta operación se vuelve clara si se recuerda la ecuación de una *B-spline* (ecuación 2) introducida anteriormente en este capítulo:

$$S(x) = \sum_{i=1}^{n} c_i B_i(x)$$

Cada función univariante que conecta dos nodos en una red KAN se representa como una combinación lineal de funciones base $B_i(x)$ donde los coeficientes c_i son los verdaderos parámetros que se ajustan durante el entrenamiento. Así, al igual que en una red tradicional se modifican pesos para minimizar el error, en una red KAN lo que se optimiza es la forma de estas funciones, mediante la actualización de sus coeficientes.

Por último, un elemento clave en la estructura de las redes KAN es el uso de un *grid* (G) o malla de puntos, que define los intervalos sobre los que se construyen los *B-splines* y constituye otro hiperparámetro del modelo. Este *grid* fija los valores de entrada x en los que cada función base $B_i(x)$ toma valores distintos de cero. Un número de *grid* más alto permite que las funciones univariantes se ajusten con mayor detalle a la entrada. Sin embargo, también implica más parámetros que entrenar, por lo que aumenta la complejidad del modelo y el riesgo de sobreajuste si no se regula adecuadamente.

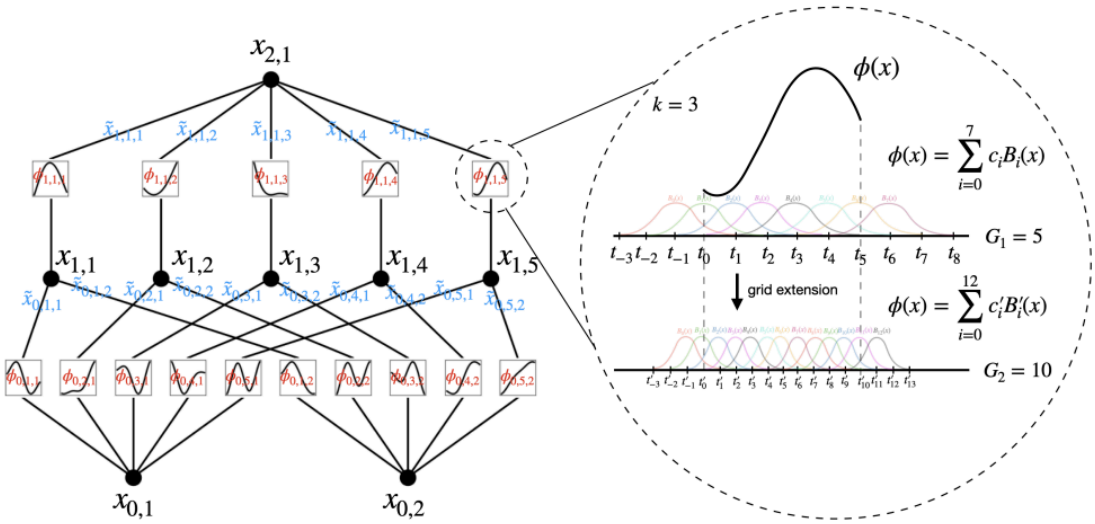

Figura 9. Representación de la arquitectura interna de una KAN, reproducido de *KAN: Kolmogorov–Arnold Networks* por Liu et al. 2024

La figura 9 ilustra de forma esquemática una red KAN profunda con múltiples capas, donde cada conexión entre nodos está asociada a una función univariante parametrizada mediante *B-splines*. En el círculo de la derecha se muestra con mayor detalle cómo se construye una de estas funciones: sobre un dominio definido por un *grid* de puntos $t_0, t_1, \ldots, t_G$, se definen funciones base $B_i(x)$ que se combinan linealmente mediante coeficientes c_i para generar la función completa $\Phi(x)$. Este *grid* determina el soporte local de cada función base y permite controlar la resolución con la que la red puede representar patrones. A medida que se incrementa el número de puntos del *grid* (como se muestra entre G=5 y G=10), la capacidad de la red para captar detalles finos aumenta.

4.1.5. Implementación práctica

Para concluir con el bloque de técnica es imprescindible entender cómo se aplican en la práctica las redes KAN a un conjunto de datos real. El proceso comienza como en cualquier tarea de aprendizaje supervisado, se parte de una base de datos estructurada con variables independientes (*inputs*) y una variable dependiente (*output*). Una red KAN se entrena para aproximar la función subyacente que relaciona ambas partes, pero en lugar de aprender pesos escalares entre neuronas, lo que se ajustan son funciones univariantes

representadas mediante *B-splines*. Estas funciones se combinan siguiendo la estructura composicional inspirada en el teorema de Kolmogorov-Arnold.

A diferencia de muchos modelos de *machine learning* considerados como "cajas negras", una de las principales ventajas de las KAN es que no solo predicen valores, sino que pueden devolver una expresión funcional simbólica que relaciona las variables de entrada con la salida. Esta propiedad las convierte en herramientas especialmente valiosas cuando se busca no solo precisión, sino también interpretabilidad y comprensión del modelo aprendido. Desde el punto de vista técnico, basta con definir el número de capas, nodos por capa y tamaño del *grid*, y luego entrenar el modelo minimizando una función de pérdida adecuada.

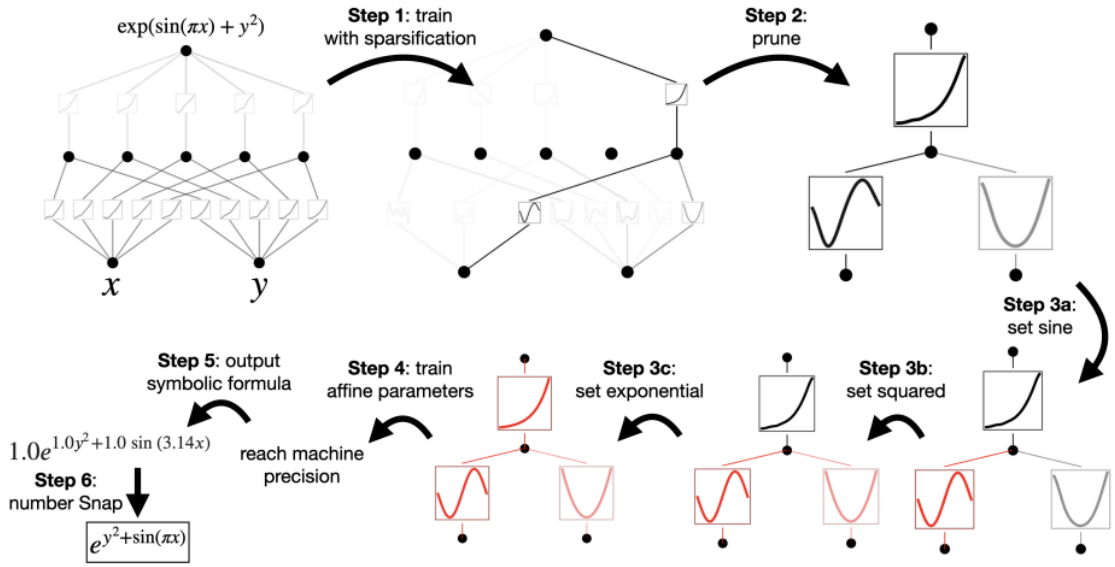

Figura 10. Proceso de entrenamiento y extracción simbólica en una red KAN, reproducido de *KAN: Kolmogorov–Arnold Networks* por Liu et al. 2024

La figura 10 ilustra un caso práctico del potencial de las redes KAN. En este ejemplo, se parte de una base de datos generada por la función que se muestra. La red KAN es entrenada sobre estos datos y, a lo largo de un proceso en seis etapas, se extrae de forma progresiva una estructura funcional interpretable. Tras un entrenamiento inicial (paso 1) y una poda de conexiones irrelevantes (paso 2), el modelo identifica automáticamente funciones conocidas como seno, cuadrado o exponencial (pasos 3a–3c). Posteriormente, se ajustan los

parámetros (paso 4) y se reconstruye una expresión simbólica equivalente (paso 5), lo que permite representar la relación aprendida en términos matemáticos explícitos (paso 6).

4.2. Marco teórico de la esperanza de vida

La esperanza de vida constituye un fenómeno complejo y multidimensional, cuya explicación no puede reducirse a un único factor. Tal como se argumentó en el capítulo 2, los determinantes que influyen en la longevidad de una población abarcan desde aspectos estructurales y socioeconómicos hasta decisiones individuales y hábitos personales. Esta comprensión se refleja en el modelo de Dahlgren y Whitehead (1991), utilizado como marco de referencia conceptual para estructurar el estado del arte. Desde este punto de vista, el fenómeno de la esperanza de vida parece compatible con la lógica de las redes KAN.

Estas redes permiten representar funciones multivariables complejas mediante la composición de funciones univariantes, lo que encaja con la hipótesis de que la longevidad puede entenderse como una combinación de factores individuales que actúan de forma no lineal y, en muchos casos, interdependiente. No obstante, como se comentó en el capítulo de restricciones, el número reducido de observaciones disponibles limita la cantidad de variables que pueden incluirse. Por ello, la selección final de variables se limita a tres determinantes escogidos siguiendo el marco conceptual, su relevancia empírica y disponibilidad en la base de datos.

5. Descripción y resultados de los experimentos

Este capítulo presenta el diseño y los resultados de los experimentos realizados para comprobar las hipótesis de la investigación. Tras haber justificado en el capítulo anterior la pertinencia teórica y metodológica del uso de redes KAN, se procede ahora a poner en práctica dicho enfoque con el fenómeno objeto del trabajo, la esperanza de vida.

En primer lugar, se describe el conjunto de datos utilizado, así como las variables seleccionadas y los objetivos concretos de los experimentos. A continuación, se detalla la implementación técnica del modelo, incluyendo las librerías empleadas, la arquitectura de la red y los criterios de selección de hiperparámetros. Finalmente, se presentan los resultados obtenidos y se analiza su coherencia con las hipótesis planteadas.

5.1. Descripción

Para el desarrollo de este estudio se ha construido una base de datos propia que incluye una variable objetivo, la esperanza de vida al nacer, junto con tres variables explicativas: el salario nacional per cápita, el gasto público en sanidad per cápita y el nivel de contaminación del aire. La unidad de análisis es el país, y los valores utilizados representan el promedio nacional de cada indicador. Todos los datos corresponden al año 2021, salvo la variable de contaminación atmosférica, que hace referencia a 2019 debido a la falta de registros más recientes y fiables en esta categoría. La información ha sido obtenida de dos fuentes oficiales: el portal World Bank Open Data, que proporciona los indicadores de esperanza de vida, salario y gasto sanitario; y la Organización Mundial de la Salud (OMS), responsable de los datos sobre niveles de contaminación del aire.

La primera variable explicativa seleccionada es el salario nacional per cápita, entendido como el ingreso promedio recibido por los trabajadores de un país en un año determinado. La elección de esta variable responde a la amplia evidencia empírica que vincula los determinantes económicos con los niveles de salud y longevidad, tal como se expuso en el capítulo 2. Entre estas contribuciones

destaca el trabajo de Preston (1975), que establece una relación sólida entre el ingreso y la esperanza de vida, mostrando que los aumentos en los niveles de renta están asociados a mejoras en los indicadores de salud a lo largo del tiempo. Además, para la comparativa entre países, como se explicó recogiendo el trabajo de Wilkinson (1997), las medidas de desigualdad económica no son significativas. Esta variable se ha preferido frente al PIB per cápita porque representa de manera más fiel el nivel de ingresos que perciben realmente las personas, y no únicamente la riqueza generada a nivel nacional.

La segunda variable explicativa incorporada en el modelo es el gasto público en sanidad per cápita, medido como la inversión pública anual destinada al sistema sanitario dividida por el número de habitantes. La inclusión de esta variable responde a la relevancia evidente que tiene la existencia de un sistema de salud adecuado para garantizar condiciones de vida que favorezcan una mayor longevidad. Sin embargo, a pesar de su aparente relación directa con la salud, diversos estudios, como el de Filmer y Pritchett (1999), han evidenciado que medir el impacto del gasto público sanitario en la esperanza de vida resulta especialmente complejo. Las dificultades se deben, en parte, a que esta variable no recoge automáticamente la calidad ni la eficiencia del sistema de salud. A pesar de estas limitaciones, se ha optado por utilizar el gasto público per cápita como una medida de esfuerzo institucional en materia sanitaria, ya que permite mantener cierta comparabilidad entre países y funciona como *proxy* razonable del compromiso estatal con la salud pública.

La tercera y última variable explicativa considerada en este estudio es la contaminación del aire, medida a través de la concentración media anual de partículas en suspensión PM2.5. Esta variable ha sido ampliamente documentada como un factor de riesgo para la salud pública, especialmente en relación con enfermedades respiratorias, cardiovasculares y diversos tipos de cáncer. Incluir esta dimensión medioambiental permite capturar no solo las condiciones estructurales y económicas de un país, sino también los factores externos que afectan a la salud de la población de forma persistente y transversal.

Tabla 2

Resumen de las variables incluidas en la base de datos

Variable	Indicador	Categoría	Fuente
Esperanza de vida	Años de vida al nacer	Variable dependiente	World Bank Open Data
Salario nacional per cápita	USD anuales por habitante (ajustado a dólares constantes de 2015)	Determinante socioeconómico	World Bank Open Data
Gasto público en sanidad per cápita	USD anuales por habitante (en dólares actuales de 2021)	Determinante institucional	World Bank Open Data
Contaminación del aire	PM2.5 (μg/m³, concentración media anual)	Determinante medioambiental	Organización Mundial de la Salud (OMS)

Elaboración propia

Para construir la base de datos, los valores de cada variable se descargaron por separado desde sus respectivos repositorios. Utilizando el código estándar de país como clave, se integraron en una única estructura. Se eligió el año 2021 por ser el más reciente con datos disponibles tanto para el salario nacional per cápita como para el gasto público en sanidad. En el caso de la contaminación del aire, se utilizó el dato de 2019, último año con cobertura global, asumiendo su relativa estabilidad a corto plazo. Tras la integración, se aplicó un filtrado para excluir los países sin registros completos, lo que redujo la muestra de 266 a 137 observaciones válidas para la fase de modelización.

A partir de esta muestra final, se diseñaron los experimentos necesarios para poner a prueba las hipótesis planteadas en el capítulo 3. El objetivo principal es evaluar si una red KAN es capaz de modelar la esperanza de vida con mayor precisión que un modelo de regresión lineal tradicional. Para ello, se comparó el rendimiento de ambos modelos utilizando dos métricas estándar: el coeficiente de determinación (R^2) y el error cuadrático medio (MSE). Estas métricas permiten

valorar tanto la capacidad de ajuste como el nivel de error cometido en las predicciones. El experimento está concebido para contrastar formalmente si la red KAN supera al modelo lineal en ambas métricas, lo cual confirmaría la hipótesis alternativa y reforzaría la utilidad de este nuevo enfoque en contextos sociales complejos como la esperanza de vida.

5.2. Implementación

Esta sección documenta el proceso técnico seguido para implementar los dos modelos utilizados en el análisis, una red KAN y un modelo de regresión lineal con efectos cuadráticos. Ambos enfoques han sido programados en lenguaje Python, utilizando el entorno colaborativo de Google Colab. Con el objetivo de realizar una comparación justa y rigurosa, se han aplicado las mismas condiciones experimentales a ambos modelos. En particular, se ha fijado la semilla aleatoria (random_state = 0) para garantizar que la división de los datos en entrenamiento, validación y test sea idéntica en ambos casos. Una mención especial merece el uso de optimización bayesiana en el caso de la red KAN, se reconoce que la optimización de hiperparámetros puede afectar la reproducibilidad exacta de los resultados.

Para evaluar el rendimiento de ambos modelos, se han seleccionado dos métricas ampliamente utilizadas: el coeficiente de determinación (R^2), que mide la capacidad explicativa del modelo, y el error cuadrático medio (MSE), que cuantifica la magnitud del error en las predicciones. Estas métricas permiten comparar los modelos tanto desde el punto de vista estadístico como desde una perspectiva práctica. A continuación se describen, de forma detallada, los procedimientos utilizados para implementar ambos modelos. El código completo empleado en la investigación se encuentra disponible en el anexo para su consulta.

5.2.1. Red KAN

Para implementar el modelo basado en redes Kolmogorov–Arnold (KAN) se utilizaron las librerías pykan, botorch, torch y scikit-learn, además de otras

auxiliares como pandas, numpy y matplotlib para la manipulación de datos y visualización. La librería pykan, desarrollada por Liu et al. (2024), permite definir arquitecturas KAN personalizadas, entrenarlas y extraer fórmulas simbólicas interpretables. Por su parte, botorch facilita la optimización bayesiana de los hiperparámetros del modelo.

El entrenamiento del modelo se estructura en tres fases: la búsqueda de hiperparámetros mediante optimización bayesiana, el ajuste del modelo final con los mejores valores encontrados y, finalmente, la poda del modelo y extracción simbólica de su fórmula. El espacio de búsqueda definido para la optimización incluye cuatro hiperparámetros clave:

- hidden: número de nodos en la capa oculta (entre 3 y 10)
- grid: número de puntos en la cuadrícula *spline* (entre 5 y 15)
- k: grado de suavidad de los *splines* (entre 11 y 21)
- λ (lamb_param): coeficiente de regularización L2, explorado en escala logarítmica entre 1e-5 y 1e-2

El resto de parámetros del modelo se fijaron durante los experimentos, entre ellos: el optimizador (Adam), el número de iteraciones de entrenamiento (steps=100), la función de pérdida (MSELoss), y la penalización de suavizado de las funciones *spline* (weight_simple=1e-3). Otros términos de regularización como lamb_l1, lamb_entropy o lamb_coef fueron establecidos en cero, de la misma forma que grid_eps.

La optimización bayesiana se realizó en dos etapas. Primero se generaron 20 combinaciones iniciales aleatorias. Posteriormente, se ejecutaron 200 iteraciones en las que se seleccionaron nuevos candidatos mediante la métrica de ganancia esperada logarítmica (LogExpectedImprovement). La evaluación de cada configuración se hizo mediante validación cruzada con cinco pliegues (KFold), para minimizar el error de validación promedio (MSE).

Una vez identificados los mejores hiperparámetros, se entrenó un modelo final sobre el conjunto combinado de entrenamiento y validación, evaluando su desempeño sobre el conjunto de test. Como paso adicional, se aplicó la función prune() integrada en la librería pykan, que simplifica la arquitectura final eliminando conexiones innecesarias. Finalmente, se extrajo la fórmula simbólica

del modelo podado utilizando una biblioteca compuesta por funciones comunes (x, x^2, log, sin, etc.), lo que permite obtener una representación interpretable de la relación entre las variables.

5.2.2. Regresión lineal

Como punto de comparación frente a la red KAN, se implementó un modelo de regresión lineal con efecto cuadrático utilizando la librería statsmodels. Este modelo actúa como referencia tradicional en tareas de predicción y permite evaluar en qué medida una arquitectura más compleja como la KAN aporta mejoras en precisión y expresividad. Con el objetivo de hacer una comparativa más fidedigna, la regresión incorpora términos cuadráticos, concatenando los valores originales escalados con sus respectivos cuadrados. Esta técnica permite capturar relaciones no lineales básicas sin abandonar la estructura lineal del modelo.

El ajuste del modelo se realizó mediante Mínimos Cuadrados Ordinarios (OLS), utilizando como conjunto de entrenamiento la partición previamente definida. Para evaluar el rendimiento sobre datos no vistos, se utilizó el conjunto de test, calculando dos métricas ya introducidas en secciones anteriores: el coeficiente de determinación R^2 y el error cuadrático medio (MSE). Finalmente, se generó una representación simbólica de la función estimada a partir de los coeficientes del modelo.

5.3. Resultados

Esta sección presenta los resultados obtenidos por ambos modelos desarrollados. Se detallan las métricas utilizadas para evaluar el rendimiento, la arquitectura resultante, así como las fórmulas simbólicas generadas por cada modelo. Finalmente, se realiza una comparativa que permite contrastar los enfoques y comprobar si se valida la hipótesis principal del estudio.

Cabe recordar que el modelo KAN incorpora un proceso de búsqueda de hiperparámetros mediante optimización bayesiana. Este procedimiento incluye

componentes aleatorios en la generación de muestras iniciales y en la exploración del espacio de parámetros, lo que implica que la arquitectura y los resultados finales pueden variar ligeramente entre ejecuciones, aunque el rendimiento general del modelo suele mantenerse dentro de un rango consistente. A pesar de que se ha fijado la semilla de aleatoriedad en pasos clave del preprocesamiento y validación cruzada, esta variabilidad es inherente al proceso de optimización utilizado.

5.3.1. Red KAN

El modelo desarrollado mediante redes KAN alcanzó unos resultados notables al predecir la esperanza de vida a partir de las variables seleccionadas. A lo largo del proceso se evaluaron tanto la versión original como su versión podada.

El modelo sin podar, que se mantuvo como referencia principal, obtuvo un coeficiente de determinación de R^2 igual a 0,846 y un error cuadrático medio de MSE igual a 5,55 sobre el conjunto de test. El conjunto óptimo de hiperparámetros, obtenido mediante un proceso de optimización bayesiana con 200 iteraciones, fue el siguiente:

- hidden = 5
- grid = 10
- k = 15
- λ (lamb_param) = $8{,}63 \cdot 10^{-4}$

La arquitectura resultante puede observarse en la figura 11, donde se visualizan los nodos, las funciones de activación adaptativas y las conexiones más relevantes:

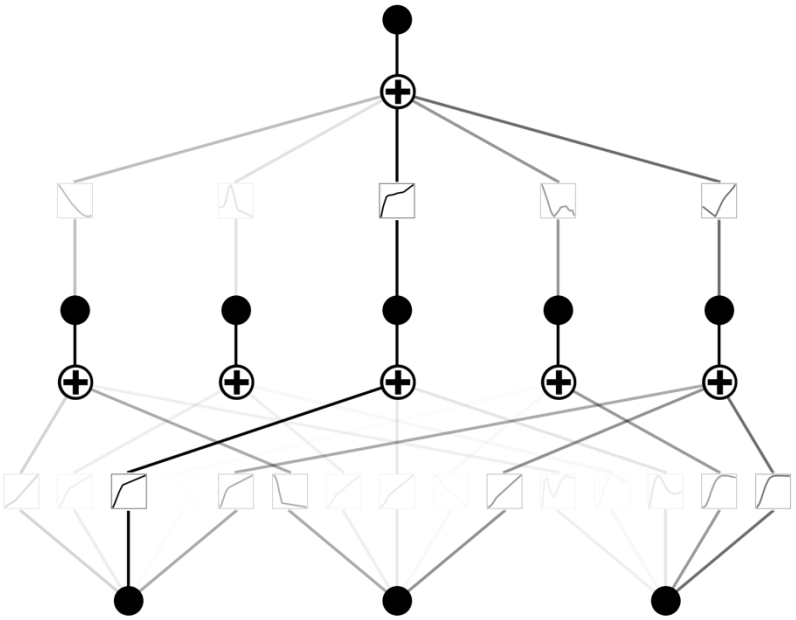

Figura 11. Arquitectura KAN, elaboración propia

Como se aprecia, la red presenta una estructura moderadamente compleja, coherente con la riqueza no lineal del fenómeno modelado. La arquitectura de la red KAN se compone de tres nodos de entrada, correspondientes a las tres variables explicativas del modelo. Estos alimentan una única capa oculta con cinco nodos intermedios, encargados de realizar combinaciones no lineales de las funciones base. Finalmente, una capa de salida con un único nodo agrega los resultados para generar la predicción final. El grosor de las líneas refleja la intensidad de las conexiones, siendo más marcadas aquellas con mayor peso en la predicción.

En este caso, la expresión resultante combina transformaciones no lineales como funciones seno, coseno, tangente hiperbólica, raíces cuadradas y exponenciales aplicadas a las variables de entrada. Esta representación ofrece una interpretación matemática del modelo, revelando de forma explícita cómo interactúan las variables explicativas en la estimación de la esperanza de vida. A pesar de su complejidad, esta fórmula representa una alternativa más transparente que otros modelos de caja negra, situándose a medio camino entre la interpretabilidad de los modelos lineales y la potencia predictiva de las redes neuronales profundas. La fórmula simbólica se muestra a continuación:

$$
\begin{aligned}
f(x_1, x_2, x_3) = \; & -3{,}0523x_3 + 0{,}0698sin(5{,}3367x_1 - 3{,}9752) \\
& + 0{,}3924sin(7{,}8014x_2 - 1{,}1917) \\
& + 20{,}9321cos(2{,}9557x_1 - 7{,}5837) \\
& - 0{,}0379cos(3{,}4720x_1 - 6{,}0041) \\
& - 0{,}0772cos(7{,}7006x_2 - 6{,}0286) \\
& + 1{,}8934cos(7{,}8270x_2 + 3{,}3874) \\
& + 0{,}0650cos(8{,}1022x_2 + 6{,}5724) \\
& + 1{,}1643cos(9{,}9688x_2 + 3{,}4000) \\
& + 14{,}9064cos\left(0{,}5055x_3 - 3{,}1766, \sqrt{\{0{,}6159 - x_1\}} - 3{,}3129\right) \\
& + 1{,}8892tanh(9{,}5999x_3 - 1{,}0480 - 5{,}6280x_1) + 63{,}5042 \\
& - 2{,}1917e
\end{aligned}
$$

Por otro lado, se evaluó también la versión podada del modelo, generada automáticamente mediante las funciones internas de la propia red KAN. Este proceso tiene como objetivo reducir la complejidad eliminando aquellos nodos y conexiones que tienen escasa o nula contribución al resultado. En la figura 12 se observa la nueva arquitectura resultante:

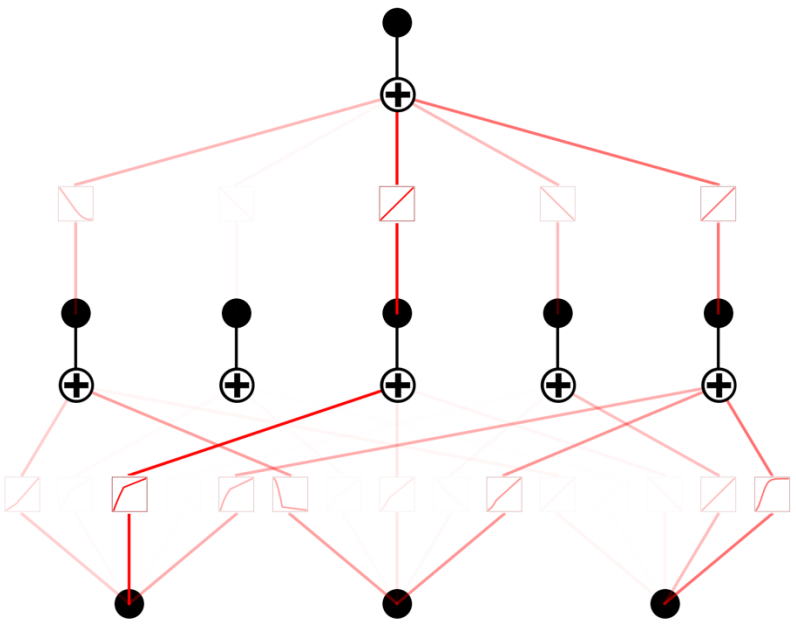

Figura 12. Modelo KAN podado, elaboración propia

En ella se aprecia una disminución del número de conexiones activas, representadas con líneas más tenues, y la desconexión de algunas funciones base. Aunque el número total de nodos por capa no varía, la poda genera un modelo funcionalmente más simple, en el que solo una fracción de los elementos participa activamente en la predicción.

En cuanto a la fórmula simbólica generada, cabe señalar que no se modifica tras la poda. La red conserva la estructura algebraica completa, pero internamente algunas contribuciones son despreciables debido a coeficientes reducidos o funciones inactivas. Por tanto, la expresión final puede mantenerse, aunque en la práctica el modelo utilice solo una parte de ella.

Esta versión simplificada obtuvo un R^2 de 0,716 y un MSE de 10,25, evidenciando una pérdida de precisión respecto al modelo completo. Por ello, y dado que la red original conserva también una buena interpretabilidad, se considera preferible optar por el modelo no podado como versión definitiva.

5.3.2. Regresión lineal

El modelo de regresión lineal con efecto cuadrático obtuvo un rendimiento algo inferior al de la red KAN. Concretamente, se alcanzó un coeficiente de determinación R^2 igual a 0,699 y un MSE de 10,84 sobre el conjunto de test. La fórmula simbólica estimada es la siguiente:

$$f(x_1, x_2, \dots, x_6) = 8.0690\,x_1 + 2.1831\,x_2 - 0.0709\,x_3 - 1.6598\,x_4 - 0.9476\,x_5 + 0.0038\,x_6 + 74.0585$$

Esta expresión destaca por su sencillez estructural, en contraste con la fórmula simbólica generada por el modelo KAN. No obstante, su capacidad para capturar relaciones no lineales complejas es limitada, lo que se refleja en un menor ajuste.

5.3.3. Comparativa

Desde un punto de vista cuantitativo, el modelo KAN presenta un rendimiento claramente superior al de la regresión lineal. En el conjunto de test, la red KAN obtuvo un coeficiente de determinación R^2 igual a 0,846 y un error cuadrático

medio MSE de 5,55. Por su parte, el modelo de regresión lineal alcanzó un R^2 de 0,699 y un MSE de 10,84. Visualmente, la diferencia también es evidente:

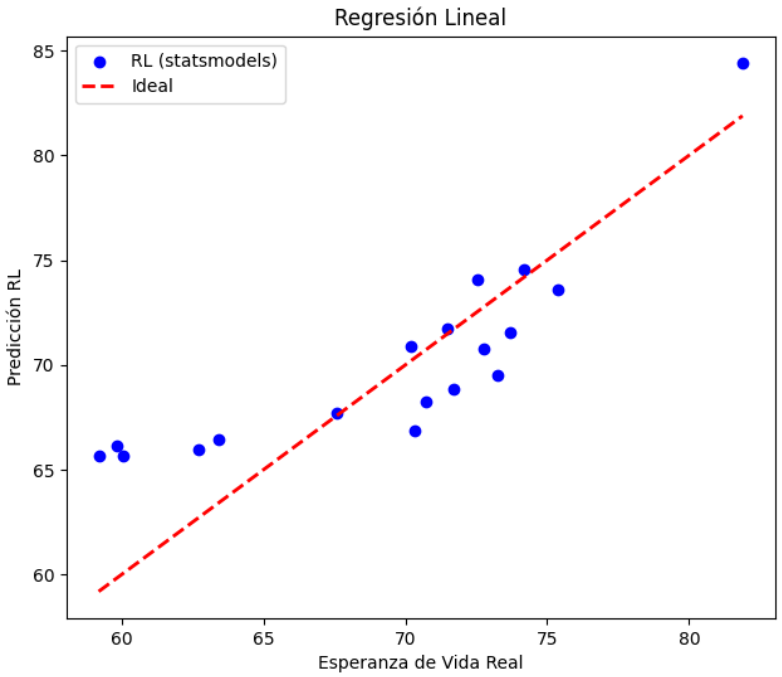

Figura 13. Predicciones regresión lineal, elaboración propia

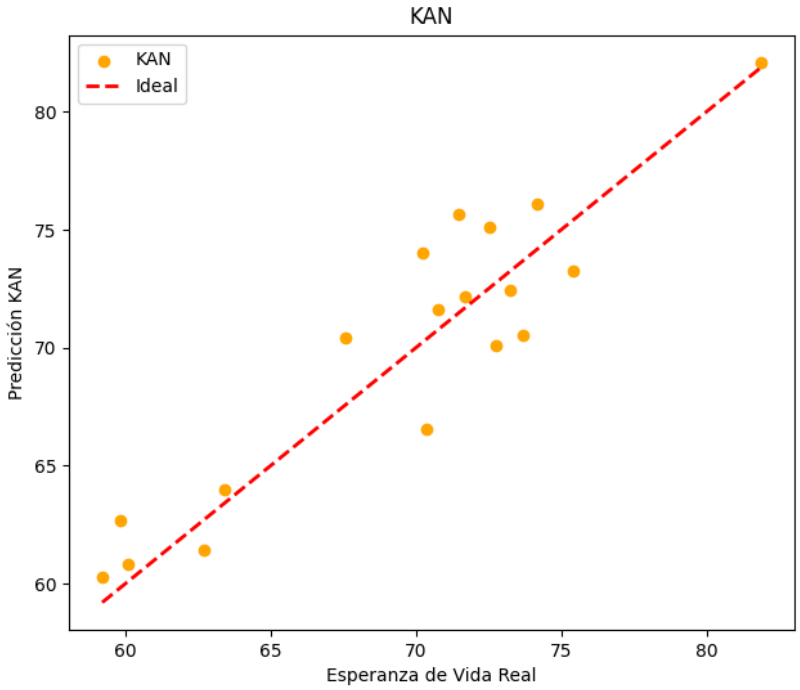

Figura 14. Predicciones KAN, elaboración propia

En las gráficas de comparación, las predicciones generadas por el modelo KAN se agrupan de manera más próxima a la diagonal ideal, lo que indica un mayor ajuste a los valores reales. En cambio, la regresión lineal muestra una mayor dispersión, especialmente en los extremos del dominio.

Desde una perspectiva interpretativa, los modelos ofrecen ventajas distintas. La regresión lineal destaca por su simplicidad y facilidad de análisis, mientras que el modelo KAN, si bien más complejo, sigue proporcionando una salida simbólica explícita. Esta capacidad para generar fórmulas interpretables en contextos no lineales es una característica diferencial de las redes KAN frente a otras arquitecturas más opacas.

En relación con las hipótesis planteadas en el capítulo 3, los resultados permiten rechazar la hipótesis nula, que sostenía que una red KAN no superaría en rendimiento a una regresión lineal. Los datos muestran que sí es posible construir un modelo simbólico basado en redes KAN que mejore las métricas de ajuste respecto a un enfoque lineal clásico. Además, el modelo resultante mantiene un nivel de interpretabilidad razonable, lo que refuerza la hipótesis alternativa del estudio: las redes KAN como una alternativa avanzada capaz de combinar precisión y transparencia. Esta capacidad no solo resulta valiosa desde el punto de vista técnico, sino que también abre la puerta a la formulación de políticas públicas más informadas y equitativas, alineadas con los Objetivos de Desarrollo Sostenible.

6. Conclusiones y resultados

Este capítulo final sintetiza los principales elementos del trabajo realizado. En primer lugar, se revisan los objetivos planteados en las etapas iniciales y se valora el grado en que han sido alcanzados. A continuación, se resumen los hallazgos más relevantes obtenidos a partir de los experimentos desarrollados, y se reflexiona sobre las implicaciones que estos tienen tanto desde el punto de vista técnico como aplicado. Finalmente, se presentan una serie de posibles líneas de investigación futura que podrían ampliar o profundizar los resultados aquí obtenidos.

6.1. Revisión de objetivos

El objetivo general de este trabajo es modelizar la esperanza de vida a partir de un conjunto de variables explicativas seleccionadas, evaluando la capacidad predictiva e interpretativa de las redes KAN en comparación con un modelo clásico de regresión lineal. A lo largo de los capítulos anteriores, se han desarrollado de forma sistemática las distintas etapas necesarias para alcanzar este propósito. En relación con los objetivos específicos planteados en el tercer capítulo:

1. Revisión de literatura: Se ha llevado a cabo una revisión detallada de los principales determinantes de la esperanza de vida que ha permitido justificar la inclusión de tres variables: el salario nacional per cápita, el gasto público en sanidad y el nivel de contaminación del aire.
2. Construcción del conjunto de datos: Se ha elaborado una base de datos que recoge información para 137 países con registros completos sobre las variables seleccionadas.
3. Implementación del modelo KAN: Se ha desarrollado un modelo basado en redes KAN, optimizado mediante un proceso de búsqueda bayesiana de hiperparámetros.
4. Desarrollo del modelo de referencia: Se implementó una regresión lineal con efecto cuadrático como modelo comparativo.

5. Evaluación comparativa de los modelos: Finalmente, se aplicaron métricas estándar (R^2 y MSE) para cuantificar el rendimiento de ambos modelos, y se extrajeron fórmulas simbólicas para valorar su grado de interpretabilidad.
6. Exploración del papel de la interpretabilidad en contextos aplicados: Se ha analizado cómo modelos simbólicos como las redes KAN pueden facilitar la comprensión de relaciones complejas y contribuir, en determinados casos, a una toma de decisiones más informada y transparente.

En conjunto, todos los objetivos establecidos han sido abordados de forma satisfactoria, permitiendo extraer conclusiones sobre la viabilidad del uso de redes KAN en contextos prácticos como el análisis de la esperanza de vida.

6.2. Resumen de hallazgos

En primer lugar, el modelo basado en redes KAN ha demostrado una superioridad clara respecto al modelo de regresión lineal, tanto en términos de ajuste como de precisión predictiva. Concretamente, el modelo KAN alcanzó un coeficiente de determinación R^2 de 0,846 y un error cuadrático medio (MSE) de 5,55, frente al R^2 de 0,699 y MSE de 10,84 obtenidos por el modelo de regresión. Esta diferencia significativa valida empíricamente la hipótesis planteada en el capítulo 3, lo que permite rechazar la hipótesis nula y confirmar que las redes KAN ofrecen un mejor rendimiento para modelizar la esperanza de vida.

Más allá del rendimiento numérico, los hallazgos permiten reflexionar sobre el equilibrio entre precisión e interpretabilidad. Si bien las redes KAN superan claramente al modelo de referencia en términos de ajuste, esta mejora se consigue a costa de una expresión simbólica considerablemente más compleja. La regresión lineal, por su parte, ofrece una fórmula más sencilla y directamente interpretable, aunque limitada en su capacidad para captar relaciones no lineales. Así, los resultados evidencian un compromiso inherente entre complejidad y precisión, que debe ser cuidadosamente valorado en función del objetivo final del análisis.

6.3. Implicaciones y aportaciones

El presente trabajo demuestra la viabilidad y el valor añadido de las redes KAN como herramienta de modelización en ciencias sociales y, más concretamente, en el análisis de fenómenos complejos como la esperanza de vida. Aunque esta arquitectura es reciente y su aplicación en este tipo de contextos todavía es escasa, los resultados obtenidos ponen de manifiesto su capacidad para capturar relaciones no lineales y generar representaciones simbólicas explícitas, lo que la sitúa como una alternativa prometedora a los modelos tradicionales.

Otra aportación destacable es la revisión estructurada de la literatura, que ha permitido identificar los principales determinantes asociados a la esperanza de vida, así como poner en evidencia aquellos aspectos menos explorados en los estudios previos. En este sentido, el trabajo no solo justifica empíricamente la selección de variables empleadas en el modelo, sino que también contribuye a ampliar el debate sobre qué factores deben tenerse en cuenta al analizar diferencias en longevidad entre países.

En conjunto, este estudio refuerza el potencial de los modelos híbridos entre precisión e interpretabilidad y propone un marco metodológico que puede ser replicado o adaptado en investigaciones futuras. Las redes KAN abren la puerta a nuevas posibilidades dentro del análisis de datos con significado estructural, especialmente en contextos donde la transparencia del modelo resulta tan importante como su capacidad predictiva. En este sentido, su aplicabilidad en ámbitos como la salud pública o la evaluación de políticas sociales adquiere un interés particular, al facilitar modelos más comprensibles y, por tanto, potencialmente más justos.

6.4. Líneas futuras de investigación

El presente trabajo ofrece múltiples vías de continuidad para investigaciones futuras que deseen profundizar en la modelización de la esperanza de vida mediante redes Kolmogorov-Arnold u otras técnicas avanzadas.

1. Incorporación de variables distintas: Aspectos como el acceso a servicios sanitarios, indicadores nutricionales, factores psicosociales o medidas de bienestar subjetivo podrían aportar una visión más integral y matizada de los determinantes de la longevidad.
2. Cambio en la unidad observacional: Trasladar el análisis a niveles más desagregados, como regiones dentro de un país, áreas urbanas, o incluso individuos, en la medida en que se disponga de los datos adecuados. Esto permitiría capturar con mayor precisión las desigualdades y particularidades locales.
3. Exploración de nuevos hiperparámetros: A pesar de haber empleado un proceso robusto de optimización bayesiana, el espacio de búsqueda se limitó a un conjunto acotado de parámetros. Futuras investigaciones podrían ampliar esta exploración incluyendo otras configuraciones, funciones base o estrategias de regularización simbólica, con el objetivo de mejorar tanto el ajuste como la interpretabilidad.
4. Comparativa con otros modelos: Finalmente, resultaría valioso comparar el rendimiento de la red KAN con otros modelos avanzados de predicción, como redes neuronales profundas o máquinas de soporte vectorial. Este análisis permitiría comprender mejor las fortalezas y limitaciones de la arquitectura KAN en relación con otras alternativas a la hora de modelizar la esperanza de vida.

A modo de cierre, este trabajo representa una contribución inicial al diálogo entre ciencia de datos, salud pública y análisis social. Aplicar herramientas matemáticas avanzadas a fenómenos humanos complejos no solo permite mejorar las predicciones, sino también descubrir nuevas relaciones que nos ayudan a comprender mejor el mundo que habitamos. Las redes KAN, con su equilibrio entre precisión e interpretabilidad, abren una ventana prometedora hacia modelos más comprensibles, accesibles y útiles para la toma de decisiones Avanzar hacia modelos más transparentes y comprensibles es también avanzar hacia decisiones más equitativas y centradas en el bienestar colectivo, una dirección que, en un mundo cada vez más interconectado y desafiante, no es solo deseable, sino necesaria.

7. Declaración de uso de herramientas de Inteligencia Artificial Generativa

Por la presente, yo, Álvaro Inchausti Vera, estudiante de Administración y Dirección de Empresas y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "REDES DE KOLMOGOROV–ARNOLD APLICADAS A LA ESPERANZA DE VIDA", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
3. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
6. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
8. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: Abril 2025

Firma:

8. Bibliografía

Bambra, C., Gibson, M., Sowden, A., Wright, K., Whitehead, M., & Petticrew, M. (2009). Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *Journal Of Epidemiology & Community Health, 64*(4), 284-291. Recuperado de https://doi.org/10.1136/jech.2008.082743

Ben-Shlomo, Y., White, I. R., & Marmot, M. (1996). Does the variation in the socioeconomic characteristics of an area affect mortality? *BMJ, 312*(7037), 1013-1014. Recuperado de https://doi.org/10.1136/bmj.312.7037.1013

Cutler, D., Deaton, A., & Lleras-Muney, A. (2006). The Determinants of Mortality. *The Journal Of Economic Perspectives*, *20*(3), 97-120. Recuperado de https://doi.org/10.1257/jep.20.3.97

Cutler, D., & Lleras-Muney, A. (2006). *Education and Health: Evaluating Theories and Evidence*. Recuperado de https://doi.org/10.3386/w12352

Dahlgren, G. &. W. M. (1991). Policies and strategies to promote social equity in health. Background document to WHO - Strategy paper for Europe. *ideas.repec.org*. Recuperado de https://lc.cx/KUOgnY

Deguen, S., & Zmirou-Navier, D. (2010). Social inequalities resulting from health risks related to ambient air quality--A European review. *European Journal Of Public Health, 20*(1), 27-35. Recuperado de https://doi.org/10.1093/eurpub/ckp220

Ezzati, M., & Riboli, E. (2013). Behavioral and Dietary Risk Factors for Noncommunicable Diseases. *New England Journal Of Medicine, 369*(10), 954-964. Recuperado de https://doi.org/10.1056/nejmra1203528

Filmer, D., & Pritchett, L. (1999). The impact of public spending on health: does money matter? *Social Science & Medicine, 49*(10), 1309-1323. Recuperado de https://doi.org/10.1016/s0277-9536(99)00150-1

Flegal, K. M., Kit, B. K., Orpana, H., & Graubard, B. I. (2013). Association of All-Cause Mortality With Overweight and Obesity Using Standard Body Mass Index Categories. *JAMA, 309*(1), 71. Recuperado de https://doi.org/10.1001/jama.2012.113905

Kawachi, I., & Kennedy, B. P. (1997). Socioeconomic determinants of health : Health and social cohesion: why care about income inequality? *BMJ, 314*(7086), 1037. Recuperado de https://doi.org/10.1136/bmj.314.7086.1037

Kruk, M. E., Gage, A. D., Joseph, N. T., Danaei, G., García-Saisó, S., & Salomon, J. A. (2018). Mortality due to low-quality health systems in the universal health coverage era: a systematic analysis of amenable deaths in 137 countries. *The Lancet, 392*(10160), 2203-2212. Recuperado de https://doi.org/10.1016/s0140-6736(18)31668-4

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., . . . Tegmark, M. (2024). KAN: Kolmogorov-Arnold Networks. *arXiv (Cornell University)*. Recuperado de https://doi.org/10.48550/arxiv.2404.19756

Mackenbach, J. P. (2012). The persistence of health inequalities in modern welfare states: The explanation of a paradox. *Social Science & Medicine, 75*(4), 761-769. Recuperado de https://doi.org/10.1016/j.socscimed.2012.02.031

Marmot, M. (2005). Social determinants of health inequalities. *The Lancet, 365*(9464), 1099-1104. Recuperado de https://doi.org/10.1016/s0140-6736(05)71146-6

Objetivos de desarrollo sostenible. Recuperado de https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/

Oeppen, J., & Vaupel, J. W. (2002). Broken Limits to Life Expectancy. Science, *296*(5570), 1029-1031. Recuperado de https://doi.org/10.1126/science.1069675

Organización Mundial de la Salud (1978). Declaración Alma-Ata. *Conferencia Internacional sobre Atención Primaria de Salud de Alma Ata*, Kazajistán.

Organización para la Cooperación y el Desarrollo Económicos (2023) *Health at a Glance 2023*. Recuperado de https://lc.cx/SpBLZC

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. *JAMA, 287*(9), 1132. Recuperado de https://doi.org/10.1001/jama.287.9.1132

Preston, S. H. (1975). The Changing Relation between Mortality and Level of Economic Development. *Population Studies, 29*(2), 231-248. Recuperado de https://doi.org/10.2307/2173509https://www.jstor.org/stable/2173509

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford, J. M., Cumming, O., Curtis, V., . . . Cairncross, S. (2014). Burden of disease from inadequate water, sanitation and hygiene in low- and middle-income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health, 19*(8), 894-905. https://doi.org/10.1111/tmi.12329

Putnam, R. D., Leonardi, R., & Nonetti, R. Y. (1994). *Making democracy work*. Recuperado de https://doi.org/10.2307/j.ctt7s8r7

Tominaga, S. (1999). Major avoidable risk factors of cancer. *Cancer Letters, 143*, S19-S23. Recuperado de https://doi.org/10.1016/s0304-3835(99)90005-8

Weston, S. P. (2013). An introduction to the mathematics and construction of splines. *Imperial*. Recuperado de https://lc.cx/o1WRVC

Wilkinson, R. G. (1997). Comment: income, inequality, and social cohesion. *American Journal Of Public Health, 87*(9), 1504-1506. Recuperado de https://doi.org/10.2105/ajph.87.9.1504

World Health Organization: WHO. (2024, 24 octubre). Contaminación del aire ambiente (exterior) y salud. Recuperado de https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

9. Anexo

9.1. Código completo red KAN

```
import math
import pandas as pd
import numpy as np
import torch
from sklearn.model_selection import train_test_split, KFold
from sklearn.preprocessing import MinMaxScaler
import matplotlib.pyplot as plt
from kan import KAN


# 1. Preprocesamiento de datos

def load_and_preprocess_data(filename="Database_v1.xlsx",
sheet="Database", target_column="Life Expectancy"):
    # Cargar datos desde Excel
    df = pd.read_excel(filename, sheet_name=sheet)
    X = df.drop(columns=[target_column, "Code"])
    y = df[target_column]

    # Dividir datos: Entrenamiento, validación y test
(random_state=0 para consistencia)
    X_train, X_temp, y_train, y_temp = train_test_split(X, y,
test_size=0.25, random_state=0)
    X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp,
test_size=0.5, random_state=0)

    # Escalar las variables a [0,1] con MinMaxScaler
    scaler = MinMaxScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_val_scaled   = scaler.transform(X_val)
    X_test_scaled  = scaler.transform(X_test)

    # Asegurar que los valores estén en el cubo unitario
    X_train_scaled = np.clip(X_train_scaled, 0, 1)
    X_val_scaled   = np.clip(X_val_scaled, 0, 1)
    X_test_scaled  = np.clip(X_test_scaled, 0, 1)

    # Convertir a tensores de PyTorch (precisión 32 bits)
    train_input = torch.tensor(X_train_scaled, dtype=torch.float32)
    train_label = torch.tensor(y_train.to_numpy()[:, None],
dtype=torch.float32)
    val_input   = torch.tensor(X_val_scaled, dtype=torch.float32)
```

```
    val_label   = torch.tensor(y_val.to_numpy()[:, None],
dtype=torch.float32)
    test_input  = torch.tensor(X_test_scaled, dtype=torch.float32)
    test_label  = torch.tensor(y_test.to_numpy()[:, None],
dtype=torch.float32)

    # Combinar entrenamiento y validación para CV
    cv_input = torch.cat([train_input, val_input], dim=0).numpy()
    cv_label = torch.cat([train_label, val_label], dim=0).numpy()
    input_dim = train_input.shape[1]
    return train_input, train_label, val_input, val_label,
test_input, test_label, cv_input, cv_label, input_dim

(train_input, train_label, val_input, val_label, test_input,
test_label, cv_input, cv_label, input_dim) =
load_and_preprocess_data()


# 2. Función para ajustar el modelo (GP)

def fit_gpytorch_model(mll, model, X, Y,
optimizer_cls=torch.optim.Adam, lr=0.1, num_steps=100):
    mll.train()
    optimizer = optimizer_cls(mll.parameters(), lr=lr)
    for i in range(num_steps):
        optimizer.zero_grad()
        output = model(X)
        loss_val = -mll(output, Y)
        loss_val = loss_val.sum() if loss_val.dim() > 0 else
loss_val
        loss_val.backward()
        optimizer.step()
        # Asegurar que el parámetro de ruido sea mayor que 0
        with torch.no_grad():

model.likelihood.noise_covar.raw_noise.data.clamp_(min=1e-4)
    return mll


# 3. Función para transformar hiperparámetros

def transform_x(x, orig_bounds):
    lower = orig_bounds[0]
    upper = orig_bounds[1]
    return lower + x * (upper - lower)


# 4. Definir la función objetivo para la optimización BO
```

```
def evaluate_objective(x, cv_input, cv_label, orig_bounds,
n_splits=5):
    x_transformed = transform_x(x, orig_bounds)
    try:
        hidden = int(round(x_transformed[0].item()))
        hidden = max(3, min(hidden, 10))
        grid = int(round(x_transformed[1].item()))
        grid = max(5, min(grid, 15))
        k_val = int(round(x_transformed[2].item()))
        k_val = max(11, min(k_val, 21))
        lamb_param = math.exp(x_transformed[3].item())
    except Exception as e:
        print("Error decodificando hiperparámetros:", e)
        return 1e6

    # Parámetros fijos para el modelo KAN
    opt_method = "Adam"
    steps = 100
    loss_fn = torch.nn.MSELoss()
    weight_simple = 1e-3
    mult_arity = 1
    noise_scale = 0.1
    scale_base_mu = 0.0
    scale_base_sigma = 1.0
    grid_eps = 0.0
    lamb_l1 = 0.0
    lamb_entropy = 0.0
    lamb_coef = 0.0
    lamb_coefdiff = 0.0

    kf = KFold(n_splits=n_splits, shuffle=True, random_state=0)
    fold_losses = []
    for train_idx, valid_idx in kf.split(cv_input):
        X_train_cv = torch.tensor(cv_input[train_idx],
dtype=torch.float32)
        y_train_cv = torch.tensor(cv_label[train_idx],
dtype=torch.float32)
        X_valid_cv = torch.tensor(cv_input[valid_idx],
dtype=torch.float32)
        y_valid_cv = torch.tensor(cv_label[valid_idx],
dtype=torch.float32)

        model = KAN(width=[X_train_cv.shape[1], hidden, 1],
                    grid=grid,
                    k=k_val,
                    mult_arity=mult_arity,
                    noise_scale=noise_scale,
                    scale_base_mu=scale_base_mu,
                    scale_base_sigma=scale_base_sigma,
```

```
                    grid_eps=grid_eps,
                    device='cpu')
        model.weight_simple = weight_simple

        try:
            results = model.fit({'train_input': X_train_cv,
                                 'train_label': y_train_cv,
                                 'test_input': X_valid_cv,
                                 'test_label': y_valid_cv},
                                opt=opt_method,
                                steps=steps,
                                loss_fn=loss_fn,
                                lamb=lamb_param,
                                lamb_l1=lamb_l1,
                                lamb_entropy=lamb_entropy,
                                lamb_coef=lamb_coef,
                                lamb_coefdiff=lamb_coefdiff)
        except Exception as e:
            print("Error durante model.fit:", e)
            return 1e6

        fold_losses.append(results["test_loss"][-1])

    avg_loss = np.mean(fold_losses)
    return avg_loss if np.isfinite(avg_loss) else 1e6


# 5. Optimización Bayesiana con BOTorch

from botorch.models import SingleTaskGP
from gpytorch.mlls import ExactMarginalLogLikelihood
from botorch.acquisition import LogExpectedImprovement
from botorch.optim import optimize_acqf

num_initial = 20
num_bo_iter = 200
X_samples = []
Y_samples = []
orig_bounds = torch.tensor([[3.0, 5.0, 11.0, math.log(1e-5)],
                            [10.0, 15.0, 21.0, math.log(1e-2)]],
dtype=torch.float64)
for _ in range(num_initial):
    x_rand = torch.empty(4, dtype=torch.float64).uniform_(0, 1)
    loss = evaluate_objective(x_rand, cv_input, cv_label,
orig_bounds)
    X_samples.append(x_rand.unsqueeze(0))
    Y_samples.append(torch.tensor([[loss]], dtype=torch.float64))
X_samples = torch.cat(X_samples, dim=0)
Y_samples = torch.cat(Y_samples, dim=0)
```

```
scaled_bounds = torch.tensor([[0.0, 0.0, 0.0, 0.0],
                              [1.0, 1.0, 1.0, 1.0]],
dtype=torch.float64)
for bo_iter in range(num_bo_iter):
    gp = SingleTaskGP(X_samples, Y_samples)
    mll = ExactMarginalLogLikelihood(gp.likelihood, gp)
    with torch.no_grad():
        gp.likelihood.noise_covar.raw_noise.data.clamp_(min=1e-4)
    fit_gpytorch_model(mll, gp, X_samples, Y_samples)
    best_f = Y_samples.min().item()
    ei = LogExpectedImprovement(model=gp, best_f=best_f)
    candidate, _ = optimize_acqf(acq_function=ei,
bounds=scaled_bounds, q=1, num_restarts=5, raw_samples=20)
    new_x = candidate.detach().squeeze(0)
    new_loss = evaluate_objective(new_x, cv_input, cv_label,
orig_bounds)
    X_samples = torch.cat([X_samples, new_x.unsqueeze(0)], dim=0)
    Y_samples = torch.cat([Y_samples, torch.tensor([[new_loss]],
dtype=torch.float64)], dim=0)

best_idx = Y_samples.argmin().item()
best_x_scaled = X_samples[best_idx]
best_loss = Y_samples[best_idx].item()
best_x = transform_x(best_x_scaled, orig_bounds)
best_hidden = int(round(best_x[0].item()))
best_grid   = int(round(best_x[1].item()))
best_k      = int(round(best_x[2].item()))
best_lamb   = math.exp(best_x[3].item())

print("\n--- Mejores Hiperparámetros ---")
print("hidden:", best_hidden, "grid:", best_grid, "k:", best_k,
"lamb_param:", best_lamb)
print("Validation Loss:", best_loss)


# 6. Entrenamiento final del modelo KAN

X_trainval = np.vstack([train_input.numpy(), val_input.numpy()])
y_trainval = pd.concat([pd.Series(train_label.numpy().flatten()),
                        pd.Series(val_label.numpy().flatten())])
trainval_input = torch.tensor(X_trainval, dtype=torch.float32)
trainval_label = torch.tensor(y_trainval.to_numpy()[:, None],
dtype=torch.float32)

final_model = KAN(width=[input_dim, best_hidden, 1],
                  grid=best_grid,
                  k=best_k,
                  mult_arity=1,
                  noise_scale=0.1,
```

```
                    scale_base_mu=0.0,
                    scale_base_sigma=1.0,
                    grid_eps=0.0,
                    device='cpu')
final_model.weight_simple = 1e-3
final_results = final_model.fit({'train_input': trainval_input,
                                 'train_label': trainval_label,
                                 'test_input': val_input,
                                 'test_label': val_label},
                                opt="Adam",
                                steps=100,
                                loss_fn=torch.nn.MSELoss(),
                                lamb=best_lamb,
                                lamb_l1=0.0,
                                lamb_entropy=0.0,
                                lamb_coef=0.0,
                                lamb_coefdiff=0.0)
print("\nResultados del modelo final:")
print(final_results)

final_model.eval()
with torch.no_grad():
    predictions = final_model(test_input)
pred_np = predictions.detach().numpy()
test_np = test_label.detach().numpy()
from sklearn.metrics import r2_score, mean_squared_error
r2 = r2_score(test_np, pred_np)
mse = mean_squared_error(test_np, pred_np)
print("\nMétricas en Test:")
print("R2:", r2, "MSE:", mse)

plt.figure(figsize=(8,6))
plt.scatter(test_np, pred_np, color='orange', label='KAN')
plt.plot([test_np.min(), test_np.max()], [test_np.min(),
test_np.max()], 'r--', lw=2, label='Ideal')
plt.xlabel("Esperanza de Vida Real")
plt.ylabel("Esperanza de Vida Predicha")
plt.title("KAN")
plt.legend()
plt.show()

final_model.plot(scale=2)
plt.figure(figsize=(10,5))
plt.plot(final_results["train_loss"], label='Pérdida en
Entrenamiento')
plt.plot(final_results["test_loss"], label='Pérdida en Validación')
plt.title('Pérdida a lo largo de Iteraciones')
plt.xlabel('Iteración')
plt.ylabel('Pérdida')
```

```
plt.legend()
plt.grid(True)
plt.show()

lib = ['x', 'x^2', 'x^3', 'x^4', 'exp', 'log', 'sqrt', 'tanh',
'sin', 'cos', 'abs']
final_model.auto_symbolic(lib=lib)
formula_kan = sp.sympify(str(final_model.symbolic_formula()[0][0]))
print("\nFÓRMULA SIMBÓLICA (KAN):")
sp.pprint(formula_kan)


# 7. Poda del modelo, evaluación y extracción simbólica

pruned_model = final_model.prune()
pruned_model.eval()
with torch.no_grad():
    pruned_predictions = pruned_model(test_input)
pruned_pred_np = pruned_predictions.detach().numpy()
r2_pruned = r2_score(test_np, pruned_pred_np)
mse_pruned = mean_squared_error(test_np, pruned_pred_np)
print("\nMétricas en Test (Modelo Poda):")
print("R2:", r2_pruned, "MSE:", mse_pruned)

plt.figure(figsize=(8,6))
plt.scatter(test_np, pruned_pred_np, color='green',
label='Predicciones Podadas')
plt.plot([test_np.min(), test_np.max()], [test_np.min(),
test_np.max()], 'r--', lw=2, label='Ideal')
plt.xlabel("Esperanza de Vida Real")
plt.ylabel("Esperanza de Vida Predicha")
plt.title("Modelo Podado: Real vs. Predicho")
plt.legend()
plt.show()

pruned_model.plot(scale=2)
plt.show()

pruned_model.auto_symbolic(lib=lib)
pruned_formula =
sp.sympify(str(pruned_model.symbolic_formula()[0][0]))
print("\nFÓRMULA SIMBÓLICA (Modelo Podado):")
sp.pprint(pruned_formula)
```

9.1. Código completo regresión lineal

```
import math
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sympy as sp
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_squared_error

# Cargar datos y definir variable objetivo
df = pd.read_excel("Database_v1.xlsx", sheet_name="Database")
target_column = "Life Expectancy"
X = df.drop(columns=[target_column, "Code"])
y = df[target_column]

# Dividir el dataset (random_state=0 en ambas divisiones)
X_train, X_temp, y_train, y_temp = train_test_split(X, y,
test_size=0.25, random_state=0)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp,
test_size=0.5, random_state=0)

# Escalado de variables predictoras
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_val_scaled   = scaler.transform(X_val)
X_test_scaled  = scaler.transform(X_test)

# Generar características cuadráticas
X_train_quad = np.concatenate([X_train_scaled, X_train_scaled**2],
axis=1)
X_val_quad   = np.concatenate([X_val_scaled, X_val_scaled**2],
axis=1)
X_test_quad  = np.concatenate([X_test_scaled, X_test_scaled**2],
axis=1)

# Ajustar modelo OLS (regresión lineal con efecto cuadrático)
X_train_quad_const = sm.add_constant(X_train_quad)
X_test_quad_const  = sm.add_constant(X_test_quad)
ols_model = sm.OLS(y_train, X_train_quad_const).fit()
y_pred_sm = ols_model.predict(X_test_quad_const)

# Métricas: R2 y MSE
r2_sm = r2_score(y_test, y_pred_sm)
```

```
mse_sm = mean_squared_error(y_test, y_pred_sm)
print("Regresión Lineal -> R2: {:.3f}, MSE: {:.3f}".format(r2_sm,
mse_sm))

# Gráfico comparativo
plt.figure(figsize=(8,6))
plt.scatter(y_test, y_pred_sm, color='blue', label='Predicciones')
plt.plot([y_test.min(), y_test.max()], [y_test.min(),
y_test.max()], 'r--', lw=2, label='Ideal')
plt.xlabel("Esperanza de Vida Real")
plt.ylabel("Esperanza de Vida Predicha")
plt.title("Regresión Lineal con Efecto Cuadrático")
plt.legend()
plt.show()

# Generar fórmula simbólica
num_features = X_train_scaled.shape[1]
symbols_linear = sp.symbols('x1:'+str(num_features+1))
symbols_quad = sp.symbols('x' + str(num_features+1) + ':' +
str(2*num_features+1))
params = ols_model.params
formula_sm = params['const']
for i, coef in enumerate(params.drop('const')):
    if i < num_features:
        formula_sm += coef * symbols_linear[i]
    else:
        formula_sm += coef * symbols_quad[i - num_features]
print("FÓRMULA SIMBÓLICA (Regresión Lineal):")
sp.pprint(formula_sm)
```