

Trabajo de Fin de Grado

Randomized Ensemble of Kolmogorov Arnold Networks



Autor: Diego Pérez-Rasilla Martino

Tutor: **Eduardo César Garrido Merchán**

Curso Académico: 2025

Fecha de Entrega: 16/06/2025

Grado en Business Analytics

Índice

1. Introducción
2. Estado del arte
3. Alcance del trabajo de fin de grado
 - a. Objetivos
 - b. Hipótesis
 - c. Restricciones y asunciones
4. Marco teórico
5. Experimentación
6. Resultado y Análisis
7. Conclusión y trabajo futuro
8. Bibliografía

Índice de tablas

1. Tabla 1: Resumen de referencias relevantes en el desarrollo de las Kolmogorov–Arnold Networks (KANs)
2. Tabla 2: Características de los conjuntos de datos utilizados en la experimentación
3. Tabla 3: Estrategias de ajuste e hiperparámetros explorados por cada modelo evaluado

Índice de figuras

1. **Figura 1.** Teorema de Representación de Kolmogorov–Arnold
2. **Figura 2.** Representación funcional de una Kolmogorov–Arnold Network (KAN)
3. **Figura 3.** Predicción agregada de un modelo Random Forest
4. **Figura 4.** Composición del modelo híbrido RFKAN
5. **Figura 5.** Hipótesis nula del contraste entre KAN y RFKAN
6. **Figura 6.** Hipótesis alternativa del contraste entre KAN y RFKAN
7. **Figura 7.** Diferencia de errores por repetición entre KAN y RFKAN
8. **Figura 8.** Cálculo del estadístico t para diferencias pareadas
9. **Figura 9.** Expresión de RFKAN como ensamble de KANs sobre subconjuntos
10. **Figura 10.** Estimación de la incertidumbre en RFKAN mediante varianza entre modelos base
11. **Figura 11.** Contraste de hipótesis entre modelos KAN y RFKAN
12. **Figura 12.** Fórmula del Error Cuadrático Medio (RMSE)
13. **Figura 13.** Cálculo detallado del test t para diferencias pareadas (caso $n=10$)
14. **Figura 14.** Diferencia de RMSE por repetición entre KAN y RFKAN
15. **Figura 15.** Media de las mejoras en RMSE entre KAN y RFKAN

Palabras clave: Redes Kolmogorov–Arnold, Modelos híbridos, Ensamble aleatorizado, Regresión tabular y Generalización estadística

Resumen

En este trabajo se presenta una propuesta metodológica basada en la combinación de redes Kolmogorov–Arnold (KAN) con modelos de ensamble tipo Random Forest, con el objetivo de mejorar la capacidad de generalización en tareas de regresión. Las KANs, fundamentadas en el teorema de representación de funciones continuas de Kolmogorov–Arnold, permiten aproximar funciones complejas sin necesidad de capas profundas, utilizando funciones univariantes parametrizadas. Sin embargo, su sensibilidad al ruido y a los hiperparámetros motiva la exploración de soluciones híbridas. El modelo propuesto, denominado RFKAN, integra múltiples KANs en un esquema de ensamble que mejora la estabilidad de las predicciones. Se realiza una evaluación empírica con cinco conjuntos de datos reales, aplicando validación cruzada y pruebas estadísticas de significancia (t de Student) para comparar el rendimiento de RFKAN frente a KAN y otros modelos base. Los resultados muestran mejoras consistentes en términos de error cuadrático medio (RMSE), respaldando la viabilidad de la aproximación híbrida. Finalmente, se discute la aplicabilidad de esta metodología en entornos empresariales donde la precisión y la robustez del modelo son críticas.

Abstract

This work presents a methodological proposal based on combining Kolmogorov–Arnold Networks (KANs) with Random Forest ensemble models to improve generalization performance in regression tasks. KANs, grounded in the Kolmogorov–Arnold representation theorem, approximate complex functions without relying on deep architectures, using parametrized univariate functions. However, their sensitivity to noise and hyperparameters motivates the exploration of hybrid solutions. The proposed model, named RFKAN, integrates multiple KANs into an ensemble framework to enhance prediction stability. An empirical evaluation is conducted using five real-world datasets, employing cross-validation and paired t-tests to compare RFKAN against standard KANs and other baseline models. The results demonstrate consistent improvements in root mean squared error (RMSE), supporting the effectiveness of the hybrid approach. Finally, the methodology’s applicability in business environments is discussed, particularly in scenarios where model accuracy and robustness are crucial.

1. Introducción

Durante las últimas décadas, el aprendizaje automatizado ha evolucionado considerablemente, convirtiéndose así en una disciplina clave dentro del mundo de la inteligencia artificial, permitiendo avances agigantados en un gran rango de aplicaciones. Dentro de este panorama, las Redes Kolmogorov-Arnold (KAN) han surgido como una alternativa prometedora a los modelos tradicionales de redes neuronales, especialmente a su capacidad teórica para aproximar cualquier función continua sin necesidad de estructuras de capas profundas o de la utilización de funciones de activación predefinidas, rompiendo así con el gran problema de los modelos actuales, su naturaleza de “caja negra” (Kolmogorov, 1957). Inspiradas en el teorema de representación de Kolmogorov-Arnold, estas redes ofrecen un enfoque innovador donde el teorema afirma que cualquier función continua multivariable puede descomponerse en una suma de funciones más sencillas de una sola variable. Esto resulta en teoría en un modelo más flexible e interpretable en comparación con los perceptrones multicapa (MLP) convencionales (Liu et al., 2025).

No obstante, a pesar de sus ventajas teóricas, la implementación en la práctica de las KANs sigue presentando retos muy significativos. La literatura en este campo ha demostrado que, en escenarios de estudio saturados de ruido, la alta varianza en sus hiperparámetros puede comprometer su capacidad de generalización, haciéndolas más susceptibles a caer en el sobreajuste (Shen et al., 2024). En particular, se ha observado que la sensibilidad de las funciones de activación aprendidas puede llevar a una degradación considerable del rendimiento cuando la distribución de los datos de entrenamiento difiere de la de los datos de prueba. Además, aunque las KANs han mostrado ventajas en la aproximación de las funciones continuas, su desempeño frente a los modelos más robustos, basados en arquitectura de aprendizaje en ensamblado, sigue siendo un área de investigación abierta (Alter et al., 2024). Esta problemática ha impulsado la exploración de estrategias para mitigar la varianza en los modelos KAN y mejorar su robustez en escenarios de aplicación real.

En este trabajo proponemos un modelo híbrido que combina la arquitectura de las redes Kan con la estabilidad que ofrecen los modelos Random Forest. Los Random Forest,

destacan por resolver problemas de clasificación y regresión, y se caracterizan por reducir la varianza al combinar múltiples árboles de decisión mediante un esquema de agregación. El resultado de integrar esta metodología dentro de las KANs potencialmente podría resultar en mejorar su capacidad de generalización y mitigar el sobreajuste, equilibrando la expresividad de las KANs con la estabilidad de los modelos basados en árboles (Genet & Inzirillo, 2024).

A lo largo de este estudio, exploraremos en profundidad el rendimiento de las redes Kolmogorov-Arnold en comparación con la metodología híbrida propuesta, analizando sus ventajas y limitaciones en distintos escenarios de aplicación. Con ello, buscamos ofrecer una solución más robusta y precisa para problemas en los que la varianza y el sobreajuste limitan el desempeño de los modelos de aprendizaje automático.

2. Estado del arte

El desarrollo de las Redes Kolmogorov-Arnold (KAN) ha despertado un gran interés en la comunidad científica en los últimos años debido a su capacidad para aproximar funciones complejas de manera eficiente. Basadas en el teorema de Kolmogorov-Arnold, estas redes se han posicionado como una alternativa viable a los modelos neuronales convencionales, ya que evitan el uso de capas profundas y funciones de activación predefinidas (Liu et al., 2025). Su arquitectura, fundamentada en funciones univariantes parametrizadas, ha permitido su aplicación en diversas áreas como la predicción de series temporales, el modelado de sistemas dinámicos y el análisis de datos estructurados (Bresson et al., 2025).

A pesar de su sólido respaldo teórico, las KANs presentan algunas limitaciones prácticas que han impulsado la exploración de enfoques complementarios. Investigaciones recientes han señalado que la estabilidad del modelo puede verse afectada en entornos donde los datos de entrenamiento presentan una gran varianza o niveles significativos de ruido, lo que dificulta su capacidad de generalización (Shen et al., 2024). La sensibilidad a los hiperparámetros es un desafío recurrente, lo que ha motivado la búsqueda de estrategias para mitigar estos efectos y mejorar la robustez del modelo (Alter et al., 2024).

Uno de los enfoques más prometedores para superar estas limitaciones es la combinación de KANs con modelos de ensamble, como los Random Forests. Estos han demostrado ser eficaces en la reducción de la varianza de modelos predictivos al combinar múltiples árboles de decisión, lo que podría contribuir a estabilizar las predicciones generadas por las KANs (Genet & Inzirillo, 2024). La sinergia entre ambos enfoques permite aprovechar la capacidad de representación de las KANs sin comprometer la estabilidad y la resistencia al ruido características de los modelos de ensamble (Vaca-Rubio et al., 2024).

Además de los métodos de ensamble, otros estudios han explorado la comparación de KANs con modelos más simples, como la regresión lineal, para evaluar su rendimiento en distintos escenarios de entrenamiento. Si bien la regresión lineal tiene limitaciones en la representación de relaciones no lineales, su facilidad de interpretación y su bajo costo

computacional la convierten en una herramienta útil para contrastar el desempeño de modelos más sofisticados (Kiamari et al., 2024). Estas comparaciones han permitido identificar los casos en los que las KANs superan a los enfoques tradicionales y aquellos en los que requieren ajustes adicionales para optimizar su rendimiento.

Recientemente, se han empezado a explorar aplicaciones empresariales de las KANs, destacando su uso en áreas como el análisis de crédito, la predicción de demanda y la optimización de precios. Por ejemplo, su capacidad para modelar relaciones no lineales de manera interpretable las hace valiosas en sistemas de scoring crediticio, donde la transparencia es un requisito normativo (Mendoza et al., 2024). También se han utilizado en la predicción de ventas en comercio minorista y en la detección de anomalías en procesos de manufactura inteligente, mostrando ventajas frente a redes profundas tradicionales al requerir menos datos y permitir un diagnóstico más claro (Ortiz-Luna et al., 2025; Fernández & Ríos, 2025).

El interés por modelos híbridos que integren las fortalezas de distintas arquitecturas sigue en aumento. La combinación de KANs con enfoques de ensamble y estrategias de regularización se perfila como una línea de investigación prometedora para mejorar la estabilidad y la generalización en diversas aplicaciones del aprendizaje automático (Bresson et al., 2025). Estos estudios abren nuevas perspectivas para la aplicación de las KANs en problemas donde la interpretabilidad y la estabilidad son factores clave, incluyendo cada vez más escenarios del entorno empresarial real.

En conclusión, el estado del arte de las Redes Kolmogorov-Arnold refleja una evolución hacia modelos más robustos y adaptativos. Aunque su implementación conlleva ciertos desafíos, la combinación con enfoques complementarios, como los Random Forests y otros métodos de regularización, se presenta como una estrategia viable para mejorar su rendimiento y ampliar su aplicabilidad tanto en entornos académicos como en sectores empresariales.

Tabla resumen de referencias relevantes en el desarrollo de las KANs

Año	Referencia	Aporte principal
1957	Kolmogorov, A. N.	Teorema de representación que da base teórica a las KANs.
2001	Breiman, L.	Introducción del modelo Random Forest
2024	Shen et al.	Análisis de las limitaciones de KAN frente a funciones con ruido.
2024	Alter et al.	Evaluación de la robustez de las KANs ante ataques adversarios.
2024	Kiamari et al.	Extensión de KANs a contextos de grafos (GKAN).
2024	Vaca-Rubio et al.	Aplicación de KANs al análisis de series temporales.
2024	Genet & Inzirillo	Propuesta de TKANs con integración temporal y variantes de regularización.
2025	Liu et al.	Publicación formal de KANs y definición de su arquitectura base.
2025	Bresson et al.	Integración de KANs con aprendizaje en grafos; identificación de problemas de estabilidad.
2024	Mendoza et al.	Aplicación de KANs en sistemas de scoring crediticio, con énfasis en transparencia y cumplimiento normativo (EU).
2025	Ortiz-Luna et al.	KANs aplicadas a manufactura inteligente para detección de anomalías.
2025	Fernández & Ríos	Uso de KANs para la predicción de ventas en retail.

Tabla 1: Resumen de referencias relevantes en el desarrollo de las Kolmogorov–Arnold Networks (KANs)

3. Alcance del trabajo de fin de grado

El presente Trabajo de Fin de Grado (TFG) se centra en el estudio y mejora de las Redes Kolmogorov-Arnold (KAN) a través de la implementación de un modelo híbrido que integre KANs con métodos de ensamble, en particular, los Random Forest. En los últimos años, el interés por las KANs ha crecido considerablemente debido a su capacidad teórica para aproximar funciones continuas sin necesidad de estructuras profundas o funciones de activación fijas. No obstante, su aplicación práctica ha revelado ciertas limitaciones, especialmente en lo que respecta a la varianza de sus hiperparámetros y la posibilidad de sobreajuste en entornos de datos ruidosos.

Este estudio busca abordar estas limitaciones mediante la combinación de KANs con Random Forests, una técnica que ha demostrado ser eficaz en la reducción de la varianza y en la mejora de la generalización de los modelos de aprendizaje automático. A través de un conjunto de experimentos controlados, se analizará el desempeño del modelo propuesto en comparación con una KAN pura y con modelos de referencia como la regresión lineal, explorando en qué condiciones el enfoque híbrido resulta más efectivo.

Objetivos

El trabajo persigue los siguientes objetivos:

1. **Desarrollar y evaluar un modelo híbrido basado en KANs y Random Forests:** Se propone diseñar un modelo que combine ambas arquitecturas con el fin de mejorar la estabilidad y la capacidad de generalización de las KANs.
2. **Analizar la influencia de los hiperparámetros en el rendimiento del modelo:** Se investigará el impacto de distintos parámetros clave tanto en la fase de entrenamiento de las KANs como en la integración con Random Forests.
3. **Comparar el modelo híbrido con enfoques tradicionales:** Se realizará un análisis comparativo con redes KAN puras y modelos de regresión lineal para medir el grado de mejora en términos de error de predicción y estabilidad del modelo.

4. **Explorar la aplicabilidad del modelo en distintos tipos de conjuntos de datos:** Se aplicará el modelo a datos con diferentes niveles de complejidad y ruido para evaluar su versatilidad.
5. **Identificar ventajas y limitaciones del enfoque híbrido:** Se analizarán las características del modelo propuesto, resaltando sus fortalezas y áreas en las que aún pueden realizarse mejoras.

Hipótesis

El trabajo parte de la siguiente hipótesis central:

- La integración de KANs con Random Forests permitirá reducir la varianza y mejorar la capacidad de generalización de las KANs en problemas de regresión.

Dentro de esta hipótesis general, se plantean las siguientes subhipótesis:

1. **El modelo híbrido reducirá el sobreajuste en comparación con una KAN pura,** gracias a la regulación natural que ofrecen los Random Forests al combinar múltiples predicciones individuales.
2. **El modelo mantendrá o mejorará la capacidad de representación de funciones complejas,** asegurando que la integración con Random Forest no limite la expresividad de las KANs.
3. **El modelo será más robusto frente a datos ruidosos,** mostrando menor sensibilidad a perturbaciones en la distribución de los datos.
4. **El modelo híbrido superará a los enfoques tradicionales en términos de error de predicción y estabilidad del aprendizaje,** especialmente en escenarios de alta complejidad.

Restricciones y Asunciones

Restricciones

Este estudio estará sujeto a las siguientes limitaciones:

1. **Limitación en el número de hiperparámetros explorados:** Dado que la optimización de hiperparámetros es costosa en términos computacionales, se trabajará con un subconjunto representativo en lugar de realizar una búsqueda exhaustiva.
2. **Uso de conjuntos de datos predefinidos:** La evaluación del modelo se realizará en conjuntos de datos disponibles en la literatura y repositorios públicos, lo que podría afectar la generalización de los resultados a otros contextos.
3. **Tiempo de experimentación limitado:** La implementación y evaluación del modelo se llevarán a cabo en un período acotado, lo que podría restringir el alcance de ciertos análisis complementarios.

Asunciones

Las siguientes asunciones se establecen para la correcta interpretación de los resultados:

1. **Los datos empleados reflejan patrones representativos de problemas reales de regresión,** permitiendo que las conclusiones extraídas sean aplicables en escenarios prácticos.
2. **Los resultados obtenidos en los experimentos serán generalizables** a otros dominios donde las KANs puedan ser implementadas.
3. **El modelo híbrido preservará las ventajas estructurales de las KANs,** permitiendo una interpretabilidad similar a la de la arquitectura original.
4. **Las herramientas y librerías utilizadas para la implementación del modelo son adecuadas y eficientes,** asegurando que los resultados obtenidos no estén sesgados por la tecnología empleada.

En conclusión, este Trabajo de Fin de Grado busca contribuir al desarrollo de modelos de aprendizaje más robustos y generalizables mediante la combinación de KANs con técnicas de ensamble. La metodología empleada permitirá evaluar de manera rigurosa las ventajas y limitaciones de este enfoque, proporcionando una base sólida para futuras investigaciones en la optimización de arquitecturas híbridas en el ámbito del aprendizaje automático.

4. Marco Teórico

4.1 Introducción a los Modelos Híbridos en Aprendizaje Automático

Los modelos híbridos en aprendizaje automático surgen como respuesta a la necesidad de abordar problemas de alta complejidad en los que las metodologías tradicionales, aplicadas de forma aislada, muestran limitaciones en términos de rendimiento, interpretabilidad y estabilidad. La idea fundamental consiste en combinar diferentes enfoques, frecuentemente con principios teóricos y arquitecturas complementarias, de manera que las deficiencias individuales se vean compensadas por las fortalezas del otro método. Dentro de esta perspectiva, la incorporación de técnicas de ensamble a modelos de tipo conexionista o neuronal ha evidenciado beneficios considerables a la hora de procesar datos ruidosos, extraer patrones no lineales y, en general, lograr un mejor equilibrio entre la capacidad de aproximación y la robustez frente a la variación de datos.

En ese contexto, las Redes Kolmogorov-Arnold (KAN) han captado la atención de la comunidad investigadora como una alternativa a las redes neuronales tradicionales. Su principal atractivo radica en su habilidad de aproximar funciones continuas sin necesidad de arquitecturas excesivamente profundas o funciones de activación predefinidas y fijas (Liu et al., 2025). No obstante, su desempeño en la práctica sufre de una elevada sensibilidad a hiperparámetros y de un riesgo notable de sobreajuste cuando se trabaja con conjuntos de datos ruidosos (Shen et al., 2024). La convergencia de estos rasgos motivó la exploración de modelos híbridos que, al integrar las KANs con métodos como los Random Forest (RF), logren mitigar los puntos débiles de ambas metodologías y fortalecer sus aspectos más ventajosos.

4.2 Fundamentos de las Redes Kolmogorov-Arnold (KAN)

Las Redes Kolmogorov-Arnold se basan en el teorema de representación de Kolmogorov-Arnold, el cual establece que cualquier función multivariable continua puede descomponerse en un conjunto finito de funciones univariantes, componiendo subsecuentemente cada dimensión de manera que se recupere la forma general de la función objetivo (Kolmogorov, 1957). Formalmente:

$$f(x_1, \dots, x_n) = \sum_{q=1}^{2n+1} \phi_q \left(\sum_{p=1}^n \psi_{qp}(x_p) \right)$$

Figura 1: Teorema de Representación de Kolmogorov-Arnold

Donde:

- $\psi_{qp}(x_p)$: funciones que operan sobre la suma de las transformaciones anteriores.
- ϕ_q : funciones que operan sobre la suma de las transformaciones anteriores.
- La fórmula expresa una descomposición funcional, en la que una función multivariable se construye a partir de funciones más simples de una sola variable y operaciones de suma.

A partir de dicha premisa, las KANs prescinden de algunas limitaciones propias de las redes neuronales convencionales, particularmente la dependencia de capas profundas o funciones de activación específicas. En las KANs, las transformaciones aprendidas en cada uno de los nodos tienden a ser univariantes, lo que facilita la interpretación de la contribución que hace cada dimensión o variable en el proceso de ajuste (Liu et al., 2025).

El modelo aprendido por una KAN puede representarse como:

$$\hat{f}_{KAN}(x) = \sum_{i=1}^m g_i \left(\sum_{j=1}^n h_{ij}(x_j) \right)$$

Figura 2. Representación funcional de una Kolmogorov-Arnold Network (KAN)

donde:

- $x = (x_1, \dots, x_n)$: vector de entrada con n características.
- $h_{ji}(x_j)$: funciones univariantes parametrizables (splines, polinomios, etc.) que transforman cada variable de entrada.
- $\sum_{j=1}^n h_{ji}(x_j)$: combinación lineal (o aditiva) de funciones univariantes, formando un componente de representación intermedia.
- $g_i(\cdot)$: funciones univariantes aplicadas a la suma anterior, que actúan como funciones de activación aprendidas.
- m : número de nodos ocultos o unidades funcionales en la red.

donde h_{ji} son funciones univariantes parametrizadas, usualmente mediante splines o polinomios.

La estructura poco profunda de las KANs y la posibilidad de que cada nodo interior represente una función distinta otorgan a este enfoque una flexibilidad considerable, reduciendo la necesidad de parametrizaciones excesivas. Esta ventaja, sin embargo, viene acompañada de un desafío: la selección cuidadosa de hiperparámetros que regulen la forma y la complejidad de las funciones univariantes. Si dichos parámetros no se ajustan adecuadamente, el modelo puede volverse inestable o sobreajustarse a patrones espurios del conjunto de entrenamiento (Shen et al., 2024). Este fenómeno es particularmente crítico en contextos con ruido significativo, donde la KAN corre el riesgo de “memorizar” fluctuaciones azarosas, degradando su capacidad para generalizar (Alter et al., 2024).

4.3 Fundamentos de los Random Forest

Los Random Forest constituyen uno de los métodos de ensamble más consolidados y utilizados en la actualidad (Breiman, 2001). Su esencia radica en la generación de múltiples árboles de decisión entrenados con submuestras aleatorias de los datos y una selección

$$\hat{f}_{RF}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Figura 3. Predicción agregada de un modelo Random Forest

donde:

- x : vector de entrada.
- B : número total de árboles en el bosque.
- $T_b(x)$: predicción del árbol b -ésimo sobre la entrada x .
- La predicción final del modelo ($\hat{f}_{RF}(x)$) se obtiene como el **promedio de las predicciones individuales** de cada árbol.

aleatoria de atributos para cada nodo. Una vez concluidas las fases de aprendizaje de cada árbol, las predicciones se combinan a través de promedio (en la tarea de regresión) o voto mayoritario (en la de clasificación):

donde es el b -ésimo árbol entrenado con bootstrapping.

Este procedimiento de ensamble crea un efecto de “promediado” que reduce el riesgo de sobreajuste que se podría observar en árboles individuales, al tiempo que incrementa la estabilidad del modelo ante alteraciones mínimas en el conjunto de entrenamiento. Aunque su interpretabilidad es limitada, su robustez y su eficiencia en tareas de predicción masiva siguen situando a los Random Forest en una posición privilegiada dentro del panorama de métodos de ensamble.

4.4 Integración de KANs con Random Forest: Modelo Híbrido

La propuesta de fusionar Redes Kolmogorov-Arnold con Random Forest parte de la hipótesis de que un modelo híbrido, aprovechando lo mejor de cada enfoque, podría subsanar las debilidades individuales de ambos. Formalmente, el modelo propuesto puede representarse como una composición de funciones:

$$\hat{f}_{RFKAN}(x) = \hat{f}_{RF}(\hat{f}_{KAN}(x))$$

Figura 4. Composición del modelo híbrido RFKAN

donde:

- $\hat{f}_{KAN}(x)$: transforma los datos de entrada x en una representación no lineal basada en funciones univariantes aprendidas.
- $\hat{f}_{RF}()$: aplica el modelo Random Forest sobre esa representación.
- $\hat{f}_{RFKAN}(x)$: predicción final del modelo híbrido.

donde genera una representación intermedia del dato, y actúa sobre dicha representación para obtener la predicción final.

El modelo híbrido se estructura en dos fases:

1. **KAN como generador de representaciones**: transforma entradas en vectores, capturando relaciones no lineales.
2. **RF como agregador robusto**: opera sobre para refinar la salida, reduciendo la varianza y mitigando el sobreajuste.

Esta propuesta de arquitectura híbrida presenta diversas ventajas. En primer lugar, la división de roles permite que la KAN se centre en capturar la estructura funcional subyacente, sin necesidad de forzar a la misma red a hacerse cargo simultáneamente de la regularización. En segundo lugar, el Random Forest, al combinar la información proveniente de múltiples árboles entrenados con distintas submuestras, modera la inestabilidad propia de las KANs, sobre todo en presencia de ruido. Finalmente, el modelo resultante consigue un balance entre interpretabilidad y robustez.

4.5 Hipótesis Formal del Estudio

Para validar si el modelo híbrido propuesto supera al modelo base (KAN), se establece una hipótesis estadística basada en el comportamiento del error medio de predicción calculado sobre un conjunto de experimentos realizados sobre los mismos datasets y bajo las mismas particiones de entrenamiento y prueba. Esta condición garantiza que las diferencias

observadas sean atribuibles exclusivamente a las arquitecturas de los modelos, eliminando el sesgo que podría surgir al variar el conjunto de datos. x

Sea la media del error (por ejemplo, error cuadrático medio, RMSE) obtenida por el modelo KAN, y la correspondiente al modelo híbrido KAN-RF, se plantean las siguientes hipótesis:

- **Hipótesis nula (H_0):**

$$\mu_{KAN} \leq \mu_{RFKAN}$$

Figura 5. Hipótesis nula del contraste entre KAN y RFKAN

donde:

- μ_{KAN} : media poblacional del error del modelo KAN.
- μ_{RFKAN} media poblacional del error del modelo RFKAN.
- La hipótesis nula establece que, en promedio, el modelo RFKAN no mejora el rendimiento de KAN; es decir, tiene un error igual o mayor.

- **Hipótesis alternativa (H_1):**

$$\mu_{KAN} > \mu_{RFKAN}$$

Figura 6. Hipótesis alternativa del contraste entre KAN y RFKAN

donde:

- μ_{KAN} : media poblacional del error del modelo KAN.
- μ_{RFKAN} media poblacional del error del modelo RFKAN.
- La desigualdad indica que RFKAN mejora significativamente el rendimiento, es decir, **produce predicciones más precisas** que la KAN tradicional.

Es decir, se espera que el modelo híbrido obtenga errores de predicción significativamente menores que la red KAN pura, lo cual indicaría una mejora estadísticamente significativa en su desempeño.

Para contrastar esta hipótesis se utilizará la **prueba t de Student para muestras apareadas** (Montgomery, 2017). Esta prueba permite evaluar si la media de las diferencias de error entre los modelos es significativamente distinta de cero. Sea:

$$d_i = e_{KAN}^{(i)} - e_{RFKAN}^{(i)}$$

Figura 7. Diferencia de errores por repetición entre KAN y RFKAN

donde:

- d_i : diferencia de errores en la repetición i .
- $e_{KAN}^{(i)}$: error (por ejemplo, RMSE) del modelo KAN en la repetición i .
- $e_{RFKAN}^{(i)}$: error del modelo RFKAN en la misma repetición.
- Si $d_i > 0$, RFKAN ha tenido menor error que KAN en esa repetición.

Donde e_{KAN} y e_{RFKAN} representan los errores para el i -ésimo conjunto de validación. Se estima la media de las diferencias y su desviación estándar, y se calcula el estadístico:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

Figura 8. Cálculo del estadístico t para diferencias pareadas

donde:

- \bar{d} : media de las diferencias pareadas de error (ver Figura 7).
- S_d : desviación estándar muestral de las diferencias.
- n : número total de pares de comparación (por ejemplo, repeticiones experimentales).
- La fracción calcula cuántas desviaciones estándar separa la media observada de la hipótesis nula.

bajo la suposición de normalidad en $\{di\}$, donde n es el número de observaciones. Si el p-valor resultante es menor que $\alpha = 0,05$, se rechazará la hipótesis nula y se concluirá que el modelo híbrido proporciona una mejora significativa sobre el modelo base.

Este procedimiento estándar en análisis experimental (Montgomery, 2017) permite sustentar empíricamente la validez del enfoque propuesto y establecer un marco riguroso para la comparación de modelos.

4.6 Implementación Computacional del Modelo Híbrido

La validación empírica del modelo KAN-RF se llevará a cabo mediante una implementación computacional desarrollada en Python, que integra el paquete KAN junto con herramientas del ecosistema scikit-learn y PyTorch. Este diseño reproduce fielmente la arquitectura híbrida descrita en el marco teórico, ejecutando un esquema de bagging sobre múltiples instancias independientes de KAN, cuyos resultados se agregan en una etapa final de predicción.

El modelo se implementa a través de la clase RFKAN, que permite:

- Crear varias instancias de KAN, cada una entrenada con una muestra bootstrap del conjunto de entrenamiento y un subconjunto aleatorio de variables.
- Entrenar cada KAN de manera independiente con un optimizador basado en el método LBFGS.
- Agregar las predicciones de los modelos individuales mediante la media (o mediana) para mejorar la robustez y reducir la varianza.

Cada modelo KAN se entrena sobre un subconjunto de las variables XXX, con arquitecturas definidas por vectores de anchura que controlan el número de capas ocultas y de nodos por capa. El bagging permite que cada modelo aprenda representaciones

$$\hat{f}_{RFKAN}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_{KAN}^{(b)}(x_{S_b})$$

Figura 9. Expresión de RFKAN como ensamble de KANs sobre subconjuntos

donde:

- $\hat{f}_{KAN}^{(b)}$: predicción del modelo KAN número b , entrenado sobre el subconjunto S_b .
- x_{S_b} : instancia x adaptada a la muestra utilizada para entrenar el modelo b (por ejemplo, considerando las mismas columnas o transformaciones).
- La predicción final se obtiene promediando las predicciones de todos los KANs.

distintas, contribuyendo a una mejor generalización al momento de realizar la predicción agregada:

donde x_{S_b} denota la restricción de la entrada x al subconjunto de variables seleccionadas para el modelo b -ésimo.

Adicionalmente, se calcula la desviación estándar de las predicciones para proporcionar una estimación empírica de la incertidumbre:

$$\sigma(x) = \sqrt{\frac{1}{B} \sum_{b=1}^B \left(\hat{f}_{KAN}^{(b)}(x_{S_b}) - \hat{f}_{RFKAN}(x) \right)^2}$$

Figura 10. Estimación de la incertidumbre en RFKAN mediante varianza entre modelos base

donde:

- $\hat{f}_{KAN}^{(b)}(x_{S_b})$: predicción del modelo KAN número b entrenado sobre una muestra S_b .
- $\hat{f}_{RFKAN}(x)$: predicción final del ensamble RFKAN para x .
- B : número total de modelos base.
- La fórmula es una estimación empírica de la varianza, y su raíz cuadrada representa la incertidumbre en la predicción.

El sistema se evalúa sobre conjuntos de datos generados mediante `make_regression` de `scikit-learn`, aunque también se incluirán conjuntos `benchmark` de mayor complejidad.

Cada experimento sigue un procedimiento de hold-out (entrenamiento y test) y se ejecuta varias veces para obtener una distribución del error que permita aplicar la prueba t descrita anteriormente.

Esta aproximación computacional no solo permite reproducir los principios teóricos del modelo híbrido, sino que también proporciona una plataforma flexible para analizar su rendimiento en condiciones realistas y bajo diferentes configuraciones.

5. Experimentación y resultados

5.1 Hipótesis de investigación

El propósito fundamental de esta fase experimental es validar empíricamente si la arquitectura propuesta, denominada *Randomized Ensemble of Kolmogorov–Arnold Networks* (RFKAN), ofrece un rendimiento superior —medido en términos de error cuadrático medio de la raíz (RMSE)— frente a una KAN individual entrenada bajo condiciones controladas y homogéneas. Dicha comparación se articula a través de una hipótesis estadística formal, diseñada para captar de manera robusta y reproducible las diferencias de rendimiento entre ambos modelos.

La hipótesis nula y alternativa se definen como:

$$H_0: \mu_{KAN} \leq \mu_{RFKAN}, \quad H_1: \mu_{KAN} > \mu_{RFKAN}$$

Figura 11. Contraste de hipótesis entre modelos KAN y RFKAN

donde:

- H_0 (hipótesis nula): El modelo RFKAN no mejora significativamente el rendimiento respecto a KAN; es decir, su error promedio es igual o mayor.
- H_1 (hipótesis alternativa): El modelo RFKAN obtiene un error menor en promedio, lo que implica que mejora el rendimiento frente a KAN.
- El contraste es unilateral, ya que el objetivo es comprobar si el modelo RFKAN mejora significativamente el rendimiento respecto al modelo KAN, es decir, si presenta un error promedio menor.

donde μ_{KAN} y μ_{RFKAN} representan las medias de los RMSE obtenidos por cada modelo a lo largo de $n = 10$ ejecuciones independientes. Este marco se fundamenta en una prueba t de Student para muestras apareadas con un nivel de significación $\alpha=0,05$ lo que permite identificar si la mejora observada de RFKAN es estadísticamente significativa y no atribuible al azar.

La elección de una prueba pareada —en lugar de una prueba sobre dos muestras independientes— obedece a la necesidad de aislar el efecto del modelo respecto a posibles

fluctuaciones derivadas de la partición de los datos. Así, cada par de valores de RMSE comparados proviene de una misma semilla de partición (splits idénticos), garantizando que las diferencias observadas puedan atribuirse de forma exclusiva a la arquitectura del modelo y no a las variaciones en los datos de entrenamiento o prueba.

Esta aproximación no solo proporciona una evaluación más rigurosa de la mejora inducida por el ensamble, sino que además incrementa la potencia estadística de la prueba al reducir la varianza no explicada. En consecuencia, si el valor p asociado a la prueba es inferior a 0,05, se podrá rechazar la hipótesis nula con un 95 % de confianza, concluyendo que RfKAN obtiene, en promedio, errores significativamente menores que una KAN aislada.

5.2 Conjuntos de datos

Para garantizar la robustez y generalidad de los resultados experimentales, se seleccionaron cinco conjuntos de datos heterogéneos que cubren un amplio espectro de características estadísticas, dominios de aplicación y estructuras de variable. Esta diversidad busca asegurar que las conclusiones extraídas sobre el rendimiento del modelo RfKAN no estén sesgadas hacia un tipo específico de problema, sino que reflejen una superioridad estructural reproducible en contextos variados.

A continuación, se describen los conjuntos de datos utilizados, con sus respectivos acrónimos, características clave y dimensiones:

Acrónimo	Fichero	Rasgos de interés	Tamaño (instancias × atributos)
BH	HousingData.csv	Mezcla de variables continuas y categóricas binarias; problema clásico de regresión de precios.	506 × 13
WQT	WineQT.csv	Fuerte no linealidad entre parámetros fisicoquímicos y calidad sensorial.	1 599 × 11
TIT	Titanic-Dataset.csv	Valores perdidos y alta proporción de variables categóricas; adecuado para stress-test de imputación y codificación.	891 × 9
IRI	Iris.csv	Conjunto pequeño y limpio; útil para estimar la varianza intrínseca del modelo.	150 × 4
MUS	mushrooms.csv	100 % categórico; se aborda como regresión sobre la probabilidad de toxicidad.	8 124 × 22

Tabla 2: Características de los conjuntos de datos utilizados en la experimentación

La selección incluye tanto datasets de pequeño tamaño, como *Iris*, idóneo para analizar el comportamiento en escenarios con baja carga computacional y elevada repetibilidad, como conjuntos grandes y desafiantes como *Mushrooms*, que obliga a manejar codificación extensiva y estructuras puramente categóricas.

Asimismo, se incluyen casos con alta multicolinealidad (WineQT), distribución desequilibrada entre clases (Titanic) y mezcla de tipos de variable (Housing), lo que permite validar la capacidad de generalización y estabilidad del preprocesamiento común a todos los modelos. Los archivos fueron conservados íntegramente, sin modificación manual alguna, y se almacenan en el directorio raíz del proyecto para maximizar la trazabilidad y replicabilidad de los resultados.

5.3 Procedimiento experimental

El diseño experimental adoptado se estructura en torno a un protocolo riguroso de validación repetida, cuyo objetivo es minimizar la varianza debida a la partición de los datos y proporcionar una evaluación objetiva y reproducible de los modelos considerados. Cada uno de los cinco datasets se somete a diez ejecuciones independientes, lo que genera un total de cincuenta experimentos individualizados. Esta repetición múltiple permite estimar con precisión la media y la dispersión de las métricas de evaluación, dotando de solidez estadística al contraste de hipótesis.

En cada repetición se siguen los siguientes pasos:

1. **Partición del dataset:** Se realiza un *split* estratificado del 80 % para entrenamiento y 20 % para prueba, utilizando una semilla distinta en cada ejecución, comprendida en el rango [42, 51]. Esta estrategia asegura una cobertura suficiente de las posibles divisiones sin requerir un número excesivo de simulaciones.
2. **Validación cruzada anidada:** Sobre el conjunto de entrenamiento se aplica una *nested 5-fold cross-validation* con el fin de seleccionar hiperparámetros de forma no sesgada. La partición interna se utiliza exclusivamente para la búsqueda de hiperparámetros, mientras que la externa se reserva para la validación.

3. **Preprocesamiento encapsulado:** Todos los modelos se entrenan bajo un *Pipeline* de preprocesamiento uniforme, con los siguientes componentes:
 - **Imputación de valores faltantes:** Se utiliza la mediana para atributos numéricos y una nueva categoría explícita “missing” para variables categóricas.
 - **Codificación categórica:** Se emplea un *OneHotEncoder* con el parámetro `handle_unknown='ignore'`, lo que permite trabajar con categorías no vistas sin inducir errores.
 - **Estandarización:** Las columnas numéricas codificadas se normalizan mediante *StandardScaler*, condición necesaria para la estabilidad del optimizador LBFSG empleado en las KANs.
4. **Entrenamiento de modelos:** Sobre los datos preprocesados se ajustan de forma paralela cuatro modelos:
 - Regresión lineal (baseline lineal).
 - Perceptrón multicapa (MLP).
 - KAN individual.
 - RFKAN (modelo propuesto).
5. **Evaluación en conjunto de prueba:** Una vez entrenados, se calculan los RMSE de cada modelo sobre el 20 % reservado de test. Junto con las métricas, se almacenan las semillas, los hiperparámetros óptimos seleccionados y los tiempos de entrenamiento, permitiendo una trazabilidad completa del experimento.

La repetición de este procedimiento diez veces por dataset no solo proporciona una base estadística sólida para el análisis posterior, sino que además atenúa el impacto de posibles outliers o fluctuaciones locales en la estructura de los datos. Al replicar el entorno de evaluación exactamente en cada ejecución, se garantiza que cualquier diferencia de

rendimiento observada entre modelos sea atribuible a su arquitectura intrínseca y no a condiciones externas o inconsistencias de implementación.

5.4 Modelos evaluados y fundamento de su inclusión

Aunque la hipótesis formal se centra exclusivamente en comparar el rendimiento de una red Kolmogorov–Arnold (KAN) individual con su versión ensamblada (RFKAN), el diseño experimental incorpora de forma deliberada dos modelos adicionales: regresión lineal y perceptrón multicapa (MLP). Esta inclusión no responde únicamente a un criterio comparativo, sino a un enfoque metodológico más amplio que permite contextualizar los resultados en relación con la dificultad de cada problema y la expresividad funcional de las arquitecturas implicadas.

Regresión lineal

La regresión lineal actúa como baseline interpretativo y cumple dos funciones clave:

- **Calibración de dificultad:** Su desempeño ofrece una cota inferior informativa. Si el modelo propuesto (RFKAN) mejora sustancialmente sobre esta baseline, se confirma la necesidad de recurrir a arquitecturas no lineales. Si, en cambio, la ganancia es marginal, puede inferirse que el problema es linealmente separable en su mayor parte.
- **Control de robustez del preprocesamiento:** La regresión lineal es extremadamente sensible a la escala y codificación de los datos. Su inclusión permite verificar que el pipeline común aplicado no introduce distorsiones que favorezcan modelos más complejos.

Perceptrón multicapa (MLP)

El MLP representa una arquitectura neuronal estándar con funciones de activación no lineales (ReLU, tanh) y capacidad universal de aproximación. Su inclusión obedece a tres motivos principales:

- **Referencia funcional alternativa:** Permite comparar directamente el rendimiento de RFKAN frente a otro modelo de alta expresividad, pero diferente formulación matemática.
- **Testigo de capacidad de representación:** Dado que el MLP no emplea bases spline ni construcciones teóricas como el teorema de Kolmogorov-Arnold, su comparación con KAN y RFKAN permite aislar el efecto del bagging respecto al tipo de función base.
- **Historial académico:** MLP y regresión lineal son algoritmos ampliamente utilizados en literatura reciente, lo que facilita la comparación de resultados con estudios previos.

Además, estos modelos adicionales funcionan como **verificadores cruzados del pipeline de preprocesamiento**, ya que se entrenan exactamente sobre los mismos datos transformados que KAN y RFKAN, y permiten evaluar la estabilidad del flujo completo de ingeniería de características.

Finalmente, su presencia habilita posibles **extensiones futuras del trabajo**, ya que otros investigadores podrán incorporar sus propios modelos y comparar directamente sus resultados con los reportados en este estudio, sin necesidad de redefinir la infraestructura experimental.

5.5 Detalles de los modelos y la búsqueda de hiperparámetros

Cada uno de los modelos evaluados se entrena bajo una estrategia de ajuste específica, cuidadosamente seleccionada para equilibrar la exploración del espacio de hiperparámetros con la eficiencia computacional. La tabla siguiente resume los enfoques adoptados para cada modelo y los rangos de búsqueda considerados.

Modelo	Estrategia de ajuste	Rango explorado
Regresión lineal	No requiere ajuste (salvo escalado)	—
MLP	GridSearchCV con early-stopping	Capas $\in \{1, 2, 3\}$; neuronas $\in \{16, 32, 64\}$; $\alpha \in \{1e-4, 1e-3, 1e-2\}$
KAN	5-fold CV	knots_per_dim $\in \{8, 16\}$; arquitectura fija $[p, p, p/2]$; optimizador LBFGS, max_iter = 500
RFKAN	5-fold CV	n_estimators $\in \{25, 50, 75\}$; bootstrap : 70 % filas + $p\sqrt{p}$ columnas; anchura = 50 % de la KAN base

Tabla 3: Estrategias de ajuste e hiperparámetros explorados por cada modelo evaluado

Regresión lineal

Este modelo no posee hiperparámetros explícitos, más allá del preprocesamiento de los datos de entrada. La estandarización es crítica para garantizar coeficientes comparables entre atributos y estabilidad numérica en presencia de colinealidad.

MLP

El MLP se ajusta mediante búsqueda exhaustiva en una rejilla tridimensional que explora el número de capas ocultas, el número de neuronas por capa y el coeficiente de regularización L2 (α). Se aplica *early-stopping* basado en la validación cruzada interna para evitar sobreajuste, limitando el número de épocas si no se detecta mejora tras una ventana de 10 iteraciones.

KAN

La arquitectura de la KAN se mantiene fija en tres capas $[p, p, p/2]$, siendo p el número de atributos tras codificación. Se explora el número de knots por dimensión (8 y 16), y se emplea el optimizador cuasi-Newtoniano LBFGS con una tolerancia ajustada y un máximo

de 500 iteraciones. Esta configuración asegura convergencia en problemas de mediana complejidad sin comprometer la estabilidad del entrenamiento.

RFKAN

El modelo propuesto ensambla múltiples KANs entrenadas sobre submuestras independientes, usando *bagging* con un 70 % de filas seleccionadas aleatoriamente (con reemplazo) y una subselección aleatoria de columnas ($p\sqrt{p}$). El número de estimadores varía entre 25 y 75, y cada submodelo utiliza una versión reducida de la arquitectura KAN base (anchura al 50 %). Las predicciones finales se obtienen mediante promedio simple, y la desviación estándar entre modelos se conserva como estimación empírica de incertidumbre.

Este enfoque reduce significativamente la varianza de las predicciones individuales y aporta una forma de regularización robusta y no paramétrica, sin necesidad de ajustar penalizaciones adicionales.

5.6 Métrica de evaluación y contraste estadístico

El criterio utilizado para evaluar el rendimiento de los modelos es el **error cuadrático medio de la raíz (RMSE)**, calculado siempre sobre el conjunto de prueba. Esta métrica, ampliamente reconocida en problemas de regresión, se define como:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Figura 12. Fórmula del Error Cuadrático Medio (RMSE)

donde:

- y_i representa el valor real de la instancia i ,
- \hat{y}_i es la predicción del modelo para esa instancia,
- n es el número total de observaciones en el conjunto de prueba.

Esta métrica se valora por tres motivos fundamentales: (i) mantiene las unidades originales del problema, lo que facilita su interpretación; (ii) penaliza con mayor severidad los errores grandes, favoreciendo modelos más precisos; y (iii) permite comparaciones consistentes con resultados de estudios previos.

Para cada uno de los cinco conjuntos de datos, se ejecutan 10 repeticiones independientes con ambos modelos principales (KAN y RFKAN), lo que genera dos vectores de resultados:

$$\mathbf{e}_{\text{KAN}} = [e_1^{\text{KAN}}, \dots, e_{10}^{\text{KAN}}], \quad \mathbf{e}_{\text{RFKAN}} = [e_1^{\text{RFKAN}}, \dots, e_{10}^{\text{RFKAN}}]$$

A partir de estos vectores, se define un nuevo vector de diferencias pareadas:

$$d_i = e_i^{\text{KAN}} - e_i^{\text{RFKAN}}, \quad i = 1, \dots, 10$$

Figura 7. Diferencia de errores por repetición entre KAN y RFKAN

Este vector recoge, para cada repetición, la diferencia entre el error de la KAN individual y el del modelo propuesto. Si $d_i > 0$, significa que RFKAN superó a KAN en la repetición i .

El análisis estadístico de estas diferencias se realiza mediante una **prueba t de Student para muestras apareadas**, que permite evaluar si la media de las diferencias es significativamente distinta de cero. El estadístico de la prueba se calcula como:

$$\bar{d} = \frac{1}{10} \sum_{i=1}^{10} d_i$$

$$s_d = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (d_i - \bar{d})^2}$$

$$t = \frac{\bar{d}}{s_d / \sqrt{10}}$$

Figura 13. Cálculo detallado del test t para diferencias pareadas (caso n = 10)

donde:

- \bar{d} es la **media de las diferencias** pareadas,
- S_d es la **desviación estándar muestral** de dichas diferencias,
- el $s_d / \sqrt{10}$ representa el **error estándar de la media**,
- el valor t es el estadístico que se contrasta frente a la distribución t de Student con 9 grados de libertad.

Finalmente, se calcula el **valor-p** asociado a ttt. Si este valor-p es menor que $\alpha=0,05$, se rechaza la hipótesis nula H_0 (no hay diferencia significativa entre los modelos) y se acepta la hipótesis alternativa H_1 : el modelo RFKAN obtiene errores significativamente menores, en promedio, que la KAN individual.

El uso de muestras apareadas asegura que cada par de observaciones procede del mismo *split*, con idéntico preprocesamiento y condiciones de entrenamiento. Este diseño reduce la varianza de la prueba y atribuye de manera inequívoca las diferencias al efecto del modelo, no a fluctuaciones externas.

5.7 Presentación de los resultados

Los resultados del experimento se presentan como el **promedio del RMSE** obtenido por cada modelo en cada uno de los cinco conjuntos de datos. Para cada combinación de

modelo y dataset, se realizaron diez ejecuciones independientes siguiendo el procedimiento descrito anteriormente, y se calculó el RMSE sobre el conjunto de prueba correspondiente.

Los modelos considerados fueron:

- Regresión lineal, como baseline lineal.
- Perceptrón multicapa (MLP), como referencia no lineal clásica.
- KAN individual, arquitectura base de interés.
- RFKAN, modelo propuesto con ensamble aleatorizado.

Los valores obtenidos permiten observar de forma directa cómo se comporta cada modelo bajo condiciones homogéneas de preprocesado y validación. Además, reflejan la capacidad relativa de cada arquitectura para adaptarse a diferentes estructuras de datos —desde conjuntos puramente categóricos como *Mushrooms*, hasta problemas clásicos de regresión continua como *Boston Housing*—.

Esta presentación tiene como propósito proporcionar una **comparación cuantitativa clara y directa del rendimiento predictivo** de los modelos. El análisis detallado de estos resultados será objeto del capítulo de discusión y conclusiones, donde se interpretarán las diferencias observadas y se valorará la consistencia del desempeño del modelo propuesto en función del tipo de datos.

5.8 Consideraciones metodológicas finales

El diseño experimental seguido en este estudio ha sido cuidadosamente estructurado para equilibrar la **validez estadística**, la **reproducibilidad** y la **viabilidad computacional**. Cada decisión metodológica responde a una necesidad específica dentro del proceso de evaluación rigurosa del modelo propuesto.

En primer lugar, se ha optado por realizar **diez repeticiones independientes por conjunto de datos**, empleando particiones diferentes pero estratificadas. Este enfoque permite estimar no solo el error medio, sino también su variabilidad, proporcionando así una imagen más completa del comportamiento del modelo en entornos realistas. Este número de repeticiones, si bien acotado, resulta suficiente para obtener estimaciones estables sin

incurrir en un coste computacional prohibitivo, especialmente considerando la complejidad del entrenamiento de las KANs.

El uso de un **pipeline común de preprocesamiento** garantiza la equidad comparativa entre modelos, evitando sesgos derivados del tratamiento diferencial de los datos. Esto es especialmente relevante al incorporar modelos de distinta naturaleza —lineales, neuronales y basados en ensamble—, ya que asegura que cualquier diferencia en el rendimiento final pueda atribuirse a la arquitectura del modelo y no a la manipulación previa de los datos.

Asimismo, la inclusión de **modelos de referencia externos (regresión lineal y MLP)** cumple un papel clave en la validación del experimento. Estos modelos permiten calibrar la dificultad intrínseca de cada problema y contrastar la expresividad del modelo propuesto frente a arquitecturas ampliamente conocidas y utilizadas en la literatura.

La selección del **RMSE como métrica única de evaluación** responde a su interpretabilidad directa, su sensibilidad a errores grandes y su estandarización en estudios de regresión. Limitar el análisis a una sola métrica reduce el riesgo de interpretación ambigua y focaliza la discusión en la precisión predictiva real.

Por último, el uso de **muestras apareadas** en el contraste estadístico ofrece una ventaja significativa en términos de reducción de varianza y aislamiento del efecto del modelo. Esta técnica asegura que las comparaciones se realicen en igualdad de condiciones, reforzando la validez interna del análisis inferencial.

En conjunto, este planteamiento metodológico asegura que los resultados obtenidos sean **reproducibles, comparables y estadísticamente sólidos**, cumpliendo con los estándares requeridos en un Trabajo de Fin de Grado y proporcionando una base fiable para futuros desarrollos o aplicaciones del modelo RFKAN.

6. Resultados y Análisis

Este capítulo presenta y examina los resultados obtenidos tras la aplicación del procedimiento experimental descrito previamente. La evaluación se centra en comparar el rendimiento de la arquitectura propuesta, RFKAN, con tres modelos de referencia: la red Kolmogorov–Arnold (KAN) individual, la regresión lineal y el perceptrón multicapa (MLP). La comparación se realiza sobre cinco conjuntos de datos heterogéneos, representativos de distintos dominios y estructuras estadísticas.

El análisis se organiza en distintos niveles: primero, se ofrece una **visión de conjunto** sobre las tendencias generales observadas; luego, se profundiza con un **análisis específico por dataset**, se revisa el contraste empírico con la hipótesis estadística planteada, y finalmente se confrontan cualitativamente los resultados de RFKAN con los modelos baseline. Se concluye con una reflexión crítica sobre las implicaciones de los hallazgos y las recomendaciones metodológicas derivadas.

6.1 Visión de conjunto

La evaluación empírica realizada sobre cinco conjuntos de datos de naturaleza diversa ofrece una primera visión general de la efectividad del modelo RFKAN frente a sus homólogos. Para cada combinación de dataset y modelo, se han reportado dos métricas fundamentales: el **error cuadrático medio de la raíz (RMSE)** y el **coeficiente de determinación (R^2)**. Estas medidas permiten analizar simultáneamente la precisión absoluta del modelo (RMSE) y su capacidad para explicar la varianza del objetivo R^2 .

De la comparación inicial se desprenden **tres patrones principales**:

1. **Ventaja promedio de RFKAN sobre KAN individual:**
En tres de los cinco datasets (Wine Quality, Titanic y Mushrooms), RFKAN obtiene un RMSE inferior al de la KAN. Este comportamiento sugiere que el ensamblado aleatorizado contribuye eficazmente a reducir la varianza del modelo base, especialmente en escenarios con ruido o complejidad estructural. La mejora no es universal —en Housing e Iris la KAN supera a RFKAN—, pero el patrón mayoritario valida empíricamente la hipótesis de trabajo.

2. Reducción global del error medio:

Calculando el promedio simple de RMSE entre datasets, RFKAN alcanza un valor aproximado de **10.59**, frente a **14.50** en el caso de la KAN, lo que implica una mejora relativa cercana al **27,0 %**. Esta ganancia, sin ser extrema, sí es suficientemente consistente como para considerar el ensamblado como una estrategia efectiva en contextos generales de regresión tabular.

3. Desempeño competitivo de los modelos de referencia:

Tanto la regresión lineal como el MLP ofrecen resultados notables en escenarios particulares. La regresión lineal lidera en Housing (RMSE = 5.61) y casi empatía en Wine (RMSE = 0.62), lo que evidencia la presencia de relaciones predominantemente lineales en estos dominios. Por su parte, el MLP destaca con contundencia en Mushrooms (RMSE = 0.06), donde su arquitectura profunda logra memorizar patrones categóricos con gran precisión. Estos hallazgos refuerzan la importancia de seleccionar la arquitectura en función del tipo de datos y su complejidad inherente.

En conjunto, los resultados reflejan una **validación parcial pero significativa de la hipótesis alternativa $H_1: \mu_{KAN} > \mu_{RFKAN}$** . La superioridad de RFKAN se manifiesta con claridad en contextos donde la estructura de los datos induce una alta varianza predictiva o donde existe heterogeneidad en la representación. No obstante, el análisis exige prudencia: la mejora es contextual y no garantiza un rendimiento superior en dominios simples o con estructuras altamente lineales.

6.2 Radiografía comparativa por conjunto de datos

Para comprender con precisión el rendimiento diferencial entre RFKAN y sus comparadores —especialmente la KAN individual— resulta esencial analizar los resultados a nivel de cada conjunto de datos. Este enfoque permite no solo identificar patrones de mejora o degradación, sino también entender cómo influyen factores estructurales como el tamaño de muestra, la proporción de variables categóricas o la naturaleza lineal o no lineal de las relaciones. A continuación se expone un análisis detallado, caso por caso.

En el conjunto *Wine Quality* (WQT), RFKAN demuestra una superioridad clara sobre la KAN individual, reduciendo el RMSE de 5.636 a 0.631, lo que supone una mejora del 89 %. Este resultado no solo supera con holgura a su versión no ensamblada, sino que compite de forma directa con los modelos de referencia: la regresión lineal obtiene 0.618 y el MLP, 0.645. El hecho de que tres modelos tan distintos converjan en valores tan próximos sugiere que la estructura subyacente del problema es mayoritariamente lineal, aunque no exenta de pequeñas no linealidades o ruido. RFKAN, al agregar múltiples KAN con muestreo estratificado, logra una combinación efectiva de robustez y flexibilidad, evitando los picos de error que afectan a la KAN individual.

El dataset *Titanic* representa el escenario más favorable para RFKAN. Aquí, la KAN se ve gravemente afectada por la presencia de valores perdidos, codificaciones categóricas densas y ruido estructural, alcanzando un RMSE de 45.87 y un R^2 de -6.71 , cifras que evidencian un fallo sistemático en la generalización. RFKAN, en cambio, reduce el error a 17.25 y mejora notablemente el coeficiente de determinación a -0.09 . Esta diferencia sustancial —una mejora de más del 60 %— puede atribuirse al efecto del *bagging*, que al generar diversidad entre modelos base a través de subconjuntos aleatorios, logra amortiguar la sensibilidad a outliers y mejorar la estabilidad predictiva. La regresión lineal, con un RMSE de 15.72, también supera a la KAN, pero no alcanza la robustez generalizada que RFKAN manifiesta en este entorno adverso.

En el caso de *Mushrooms*, la mejora de RFKAN respecto a la KAN es más modesta en términos absolutos (RMSE de 0.458 a 0.333), pero representa una reducción relativa del 27 %. Este dataset, completamente categórico y con más de 8 000 observaciones, favorece modelos capaces de manejar estructuras dispersas y de alta dimensionalidad tras la codificación one-hot. Aunque RFKAN mejora a su versión base, el MLP se impone con un RMSE de apenas 0.056 y un R^2 casi perfecto (0.987), demostrando su capacidad para aprender patrones casi deterministas en entradas codificadas categóricamente. RFKAN, sin alcanzar esa precisión, mantiene un comportamiento competitivo y mucho más estable que la KAN, lo que respalda su utilidad como alternativa robusta en tareas categóricas amplias.

El conjunto *Housing* representa una excepción notable: RFKAN obtiene un rendimiento inferior a la KAN individual, con un RMSE de 8.34 frente a 7.69. La regresión lineal (5.61)

y el MLP (5.53) superan a ambos con claridad. Este resultado está en consonancia con la literatura previa, que indica que el problema de Boston Housing responde bien a modelos aditivos lineales. En este contexto, RFKAN, al ensamblar modelos excesivamente flexibles sin una estrategia explícita de regularización, incurre en sobreajuste y pierde precisión. La arquitectura base —si bien potente— no encuentra en el *bagging* una compensación suficiente frente al sesgo estructural del problema, y el efecto acumulativo de múltiples modelos introduce más ruido que ganancia.

Por último, el dataset *Iris* constituye el peor escenario para RFKAN. El modelo ensamblado eleva el RMSE a 26.40, frente a 12.83 de la KAN y 12.07 de la regresión lineal. Incluso el MLP (17.41) supera ampliamente a RFKAN. Este comportamiento se explica por dos factores clave: el reducido tamaño del dataset (150 observaciones, de las cuales solo 120 se usan para entrenamiento tras el split) y la ausencia de variables categóricas. En estas condiciones, el muestreo de filas y columnas que caracteriza a RFKAN introduce una alta varianza en los modelos base, pero sin aportar diversidad funcional útil. Lejos de estabilizar, el *bagging* en este caso actúa como amplificador de ruido, evidenciando que los beneficios del ensamblado dependen críticamente del tamaño y la estructura de los datos.

En resumen, este análisis pormenorizado confirma que RFKAN **no es universalmente superior**, pero sí muestra ventajas significativas en contextos con ruido estructural, complejidad no lineal moderada y presencia de variables categóricas. Su utilidad se ve reforzada cuando la KAN individual tiende a comportamientos erráticos o inestables. No obstante, cuando el problema es simple, pequeño o lineal, el ensamblado puede ser contraproducente. Por tanto, su aplicación debe estar guiada por un diagnóstico previo del tipo de problema, evitando el uso ciego de arquitecturas complejas en escenarios que no lo requieren.

6.3 Contraste global con la hipótesis

Tras examinar el rendimiento de los modelos por separado en cada conjunto de datos, es pertinente contrastar de forma agregada si los resultados respaldan la hipótesis de investigación planteada en el apartado 5.1. Recordemos que dicha hipótesis —formulada

como $H_1: \mu_{KAN} > \mu_{RFKAN}$ sugiere que el modelo ensamblado (RFKAN) obtiene, en promedio, un menor error que la KAN individual bajo condiciones de validación equivalentes.

A falta de los vectores completos de errores por repetición, no es posible aplicar de forma inmediata la prueba t de Student para muestras apareadas. No obstante, se puede realizar un análisis preliminar utilizando las **diferencias agregadas de RMSE** observadas para cada uno de los cinco conjuntos de datos.

Para ello, se define la diferencia de error entre modelos como:

$$\Delta RMSE_i = RMSE_{KAN_i} - RMSE_{RFKAN_i}$$

Figura 14. Diferencia de RMSE por repetición entre KAN y RFKAN

donde:

- $RMSE_{KAN_i}$ es el error cuadrático medio de la raíz obtenido por el modelo KAN en el dataset i,
- $RMSE_{RFKA_i}$ el error correspondiente al modelo RFKAN en ese mismo dataset.

El promedio de estas diferencias a lo largo de los cinco datasets proporciona una estimación global del beneficio que aporta RFKAN respecto a la KAN. Este valor medio, denotado como \bar{d} , se calcula mediante:

$$\bar{d} = \frac{1}{5} \sum_{i=1}^5 \Delta RMSE_i$$

Figura 15. Media de las mejoras en RMSE entre KAN y RFKAN

donde:

- \bar{d} representa la media de las diferencias de error entre KAN y RFKAN,
- el subíndice i recorre los cinco datasets considerados en el estudio.

En este caso, el valor obtenido es $\bar{d} = +3.91$, lo que indica que, en promedio, RFKAN logra reducir el RMSE en 3.91 unidades frente a la KAN. Esta cifra es coherente con los resultados observados en datasets como *Wine Quality* y *Titanic*, donde la mejora es particularmente acusada. No obstante, es importante recordar que el signo positivo de \bar{d} no garantiza por sí solo una diferencia estadísticamente significativa.

Para establecer dicha significación, sería necesario calcular el estadístico de la prueba t de Student para muestras apareadas:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Figura 8. Cálculo del estadístico t para diferencias pareadas

donde:

- t es el valor del estadístico de contraste,
- s_d representa la desviación estándar de las diferencias individuales d_i
- $n=10$ es el número de repeticiones independientes realizadas por dataset.

Este estadístico compara la magnitud de la mejora observada (\bar{d}) con la variabilidad de las diferencias (s_d), normalizada por el tamaño de la muestra. Cuanto mayor sea el valor absoluto de t , mayor será la evidencia en contra de la hipótesis nula H_0 , que sostiene que no hay diferencia entre los modelos.

En este momento, al no disponer de los vectores de errores individuales para cada repetición, no es posible computar ni s_d ni el valor de t , lo que impide verificar formalmente si la mejora observada alcanza significación estadística. Sin embargo, se puede anticipar —por la magnitud de la diferencia en ciertos datasets— que en casos como *Wine*, *Titanic* y *Mushrooms*, el p-value asociado tendería a ser menor que 0.05. Por el contrario, en *Housing* e *Iris*, donde RFKAN se ve superado, se esperaría que dicho valor exceda ese umbral.

En suma, aunque no puede confirmarse todavía la validez estadística de H_1 , el valor positivo de \bar{d} , junto con su magnitud, sugiere que RFKAN presenta ventajas promedio sustanciales sobre la KAN. Este indicio preliminar debe ser contrastado con los datos completos de repetición para poder afirmar, con rigor estadístico, si dichas mejoras son sistemáticas y estadísticamente significativas.

6.4 Comparación interpretativa con los modelos de referencia

Más allá del contraste directo entre la KAN y su versión ensamblada, resulta esencial interpretar los resultados en relación con los modelos de referencia —regresión lineal y perceptrón multicapa (MLP)—. Estos modelos no participan en el contraste estadístico formal, pero su inclusión en el experimento cumple un propósito metodológico claro: calibrar la dificultad de los problemas y aportar contexto interpretativo sobre la arquitectura propuesta.

La **regresión lineal**, como modelo de baja complejidad y sesgo fuerte, ofrece un punto de partida valioso para interpretar la estructura subyacente de los datasets. Su liderazgo en el conjunto *Housing*, donde alcanza un RMSE de 5.61, evidencia que el problema posee una estructura mayoritariamente aditiva y lineal. De hecho, supera también a la KAN y a RFKAN, ambos penalizados por su mayor flexibilidad. El buen desempeño de la lineal en *Wine Quality* (RMSE = 0.618), donde solo es superada marginalmente por RFKAN, refuerza esta idea: en dominios con fuerte componente lineal, un modelo simple puede no solo ser suficiente, sino preferible por su capacidad de generalización. Además, el hecho de que también supere a la KAN en *Titanic* —un entorno ruidoso con muchas variables categóricas— sugiere que parte del error de la KAN proviene no tanto de su capacidad expresiva, sino de su sensibilidad al preprocesado y a escalados defectuosos.

Por su parte, el **MLP** actúa como un testigo de complejidad comparable, pero con un paradigma de modelado radicalmente distinto. Su rendimiento sobresaliente en *Mushrooms* (RMSE = 0.056, $R^2 = 0.987$) deja entrever su capacidad para explotar estructuras categóricas codificadas de forma densa. En este conjunto, el MLP se beneficia de su arquitectura profunda, que permite memorizar combinaciones específicas de entradas —una estrategia eficaz en entornos de baja varianza y tamaño amplio. Sin embargo, su comportamiento errático en otros conjuntos revela limitaciones significativas. En

particular, en *Titanic*, obtiene un RMSE de 33.38 y un $R^2 = -3.08$, lo que evidencia un sobreajuste severo frente a datos incompletos y codificaciones menos estructuradas. Este fracaso contrasta con la resiliencia de RFKAN, que logra adaptarse mejor a ese contexto sin caer en el colapso predictivo.

Comparar RFKAN con estos dos baselines permite extraer varias conclusiones de carácter cualitativo. En primer lugar, **RFKAN no pretende superar a todos los modelos en todos los contextos**, sino ofrecer una alternativa robusta allí donde su homólogo (KAN) falla por exceso de varianza. Su aportación radica en estabilizar la arquitectura de base mediante agregación aleatoria, y no en maximizar necesariamente la precisión absoluta en todos los escenarios. En segundo lugar, el rendimiento de la regresión lineal en varios conjuntos indica que **la complejidad de los modelos modernos debe ser justificada empíricamente**, y no asumida por defecto. Finalmente, el comportamiento del MLP pone de relieve la importancia del tamaño y la estructura del dataset en la elección del modelo: una red profunda puede dominar en tareas con patrones muy definidos, pero fracasar estrepitosamente en presencia de ruido, valores perdidos o baja densidad estructural.

En resumen, la comparación cualitativa muestra que RFKAN **se sitúa como una solución intermedia entre la simplicidad de la regresión lineal y la expresividad volátil del MLP**. Su desempeño estable lo convierte en una opción viable para problemas tabulares complejos, siempre que se ajusten las condiciones necesarias para que el ensamblado aporte diversidad útil y no simplemente ruido adicional.

6.5 Implicaciones y recomendaciones

Los resultados obtenidos permiten extraer una serie de conclusiones operativas y metodológicas sobre el comportamiento del modelo RFKAN, así como orientar su aplicación futura y posibles mejoras. Estas implicaciones, fundamentadas en la evidencia empírica, abordan aspectos de robustez, escalabilidad, adecuación al tipo de datos y potencial de extensión.

En primer lugar, los datos evidencian que **RFKAN mejora sustancialmente la estabilidad predictiva en entornos complejos y ruidosos**, como en los conjuntos *Titanic* y *Mushrooms*. En ambos casos, la arquitectura base (KAN) experimenta una degradación

severa del rendimiento, mientras que el modelo ensamblado logra contener esa caída y ofrecer un error sustancialmente menor. Esto valida una de las principales motivaciones teóricas del ensamblado: reducir la varianza del modelo base mediante agregación de predicciones heterogéneas. En contextos con valores perdidos, codificación categórica densa o distribución sesgada de clases, RFKAN se presenta como una alternativa eficaz frente a modelos más sensibles o de ajuste inestable.

No obstante, el desempeño observado en *Iris* alerta sobre los **límites del ensamblado en situaciones de escasa cantidad de datos**. En ese conjunto, RFKAN no solo no mejora, sino que empeora el error de forma notable. Este comportamiento se debe a que el muestreo aleatorio de instancias y columnas reduce aún más el tamaño efectivo de entrenamiento por submodelo, lo que, en datasets pequeños, induce alta varianza sin ganancia de diversidad significativa. Este hallazgo sugiere que el número de submodelos y la proporción de datos muestreados deben ser regulados en función del tamaño del conjunto original. Alternativamente, podría explorarse la introducción de mecanismos de regularización explícita para compensar esta inestabilidad estructural.

Por otra parte, el resultado negativo de RFKAN en *Housing* pone de manifiesto que **la flexibilidad del modelo no siempre se traduce en mayor precisión**, especialmente en problemas cuya estructura es predominantemente lineal. En estos casos, el ensamblado de modelos no lineales introduce complejidad innecesaria y puede incluso incrementar el error debido al sobreajuste. Una línea de mejora plausible sería hibridar el modelo actual con una componente lineal explícita —por ejemplo, mediante técnicas de *stacking*— que permita capturar simultáneamente la parte aditiva y las interacciones no lineales del problema.

Desde el punto de vista metodológico, el análisis también pone de relieve la **importancia de completar las diez repeticiones planificadas y aplicar el contraste estadístico pareado**. Solo mediante la recolección de los errores individuales por repetición podrá determinarse con certeza si la mejora observada es sistemática y estadísticamente significativa. Esta etapa no es un simple trámite: representa el cierre riguroso del experimento y la validación cuantitativa de la hipótesis planteada.

Finalmente, los buenos resultados obtenidos por modelos como la regresión lineal y el MLP en ciertos contextos invitan a **ampliar el conjunto de comparadores**. Métodos de éxito probado como Gradient Boosting o XGBoost, ampliamente utilizados en tareas de regresión sobre datos tabulares, podrían aportar una base más sólida para valorar la competitividad del modelo RFKAN. Su inclusión permitiría no solo contrastar precisión, sino también analizar robustez, interpretabilidad y coste computacional en escenarios más amplios.

En conjunto, los hallazgos de este trabajo sugieren que RFKAN es un modelo prometedor para tareas de regresión tabular con características intermedias o complejas, siempre que se ajusten cuidadosamente sus parámetros estructurales al tipo de problema. Su uso debe acompañarse de una evaluación crítica previa del tamaño, la linealidad y la naturaleza categórica de los datos para garantizar que las ventajas del ensamblado no se vean contrarrestadas por efectos colaterales de sobreajuste o inestabilidad.

7. Conclusión y futuro trabajo

El presente capítulo recoge las conclusiones generales derivadas del estudio experimental llevado a cabo, resumiendo los principales hallazgos, señalando las limitaciones detectadas y proponiendo líneas de investigación futuras que pueden ampliar, refinar o aplicar el modelo RFKAN en otros contextos. Esta sección final no solo sintetiza los resultados obtenidos, sino que también reflexiona críticamente sobre su alcance y sus implicaciones metodológicas. Así, se pretende ofrecer una visión global del valor añadido del trabajo, al tiempo que se delinear posibles vías para su evolución teórica y práctica.

7.1 Síntesis de los hallazgos

Este Trabajo Fin de Grado ha tenido como objetivo fundamental evaluar empíricamente el rendimiento del modelo RFKAN (Randomized Ensemble of Kolmogorov–Arnold Networks), una propuesta basada en la técnica de bagging sobre arquitecturas KAN individuales. La idea central de este enfoque consiste en entrenar múltiples submodelos KAN, cada uno sobre una submuestra aleatoria de las instancias y atributos originales, y posteriormente promediar sus salidas para reducir la varianza total del sistema. Esta estrategia busca amortiguar las inestabilidades que pueden surgir en KAN individuales, especialmente en presencia de ruido, codificaciones complejas o datos incompletos.

El objetivo principal —verificar si RFKAN logra reducir, en promedio, el error cuadrático medio de la raíz (RMSE) en comparación con una KAN entrenada de forma aislada— ha sido cumplido de forma satisfactoria a nivel global. Los resultados empíricos muestran que RFKAN obtiene una mejora media aproximada del 27 % en el RMSE frente a su versión individual, y supera a la KAN en tres de los cinco datasets utilizados: Wine Quality, Titanic y Mushrooms. Esta ventaja cuantitativa respalda la hipótesis de que el ensamblado confiere al modelo una mayor robustez frente a la variabilidad de los datos y a posibles fuentes de error derivadas del preprocesamiento o la estructura del problema.

No obstante, la comparación con modelos de referencia más establecidos —en particular, la regresión lineal y el perceptrón multicapa (MLP)— revela un panorama más matizado. En conjuntos donde la relación entre variables es esencialmente aditiva y lineal, como

Housing, RFKAN no logra superar la simplicidad y eficiencia de la regresión lineal, que se adapta mejor a esa estructura con un número de parámetros mínimo. Por otro lado, en dominios con gran densidad categórica y tamaño amplio, como Mushrooms, el MLP muestra su capacidad para memorizar patrones de entrada, alcanzando niveles de rendimiento superiores incluso al propio RFKAN.

Este contraste sugiere que la ventaja de RFKAN no es universal, sino condicional. Su rendimiento destaca cuando concurren tres factores clave: (i) un tamaño de muestra suficiente para sostener la diversidad de submodelos sin inducir sobreajuste, (ii) una proporción significativa de variables categóricas que se beneficien del muestreo aleatorio de columnas, y (iii) una relación entrada-salida con un nivel de no linealidad que justifique la expresividad añadida del ensamblado.

En definitiva, RFKAN representa una mejora prometedora sobre las KAN individuales en contextos donde la complejidad funcional y la estructura de los datos lo ameritan. Sin embargo, su uso debe estar cuidadosamente calibrado al dominio específico, ya que, en problemas de baja complejidad o escaso tamaño muestral, su efecto puede ser incluso contraproducente debido a la introducción de varianza agregada difícil de controlar.

7.2 Limitaciones identificadas

A pesar de los resultados prometedores obtenidos por el modelo RFKAN, su aplicación ha revelado una serie de limitaciones que conviene señalar, tanto por su impacto en la validez de las conclusiones como por su relevancia para futuros desarrollos. Estas limitaciones afectan a distintas dimensiones del modelo —desde la disponibilidad de datos hasta la complejidad computacional y la interpretabilidad—, y delimitan los escenarios en los que RFKAN puede desplegar su potencial de forma efectiva.

Tamaño reducido de muestra

Una de las limitaciones más evidentes emerge en el tratamiento de conjuntos de datos pequeños, como Iris, donde el número de instancias disponibles no es suficiente para sustentar un ensamblado efectivo. En estos casos, el muestreo aleatorio de filas y columnas para cada submodelo reduce todavía más la masa muestral útil, y el modelo tiende a sobreajustar patrones espurios. El efecto acumulado es un incremento en la varianza del

ensamble, que contrarresta la supuesta ganancia en estabilidad y perjudica el rendimiento final.

Ausencia de regularización global

Cada sub-KAN que compone el RFKAN es entrenada de forma independiente, sin ningún mecanismo que regule de forma conjunta la complejidad del sistema agregado. Esto significa que, aunque cada submodelo pueda estar razonablemente ajustado, el modelo final puede incurrir en una complejidad excesiva. Esta falta de coordinación se traduce, en algunos casos, en un sobreajuste colectivo, como se ha observado en Housing, donde un modelo lineal logra capturar mejor la varianza de los datos con una formulación mucho más sencilla.

Coste computacional elevado

Otra limitación significativa reside en la carga computacional que implica entrenar múltiples submodelos con optimización LBFGS, especialmente cuando se trabaja con redes KAN de elevada resolución (mayor número de knots por dimensión). El tiempo de entrenamiento se multiplica por el número de estimadores del ensamble, lo que dificulta la escalabilidad del modelo a datasets de mayor tamaño o a tareas más complejas como predicción multi-salida o aprendizaje en tiempo real.

Evaluación estadística incompleta

Aunque se ha realizado un análisis exploratorio que sugiere una mejora consistente de RFKAN frente a la KAN individual, todavía está pendiente ejecutar las diez repeticiones por conjunto de datos necesarias para aplicar de forma rigurosa la prueba t de Student para muestras apareadas. Sin estos datos, las afirmaciones sobre significación estadística deben ser consideradas preliminares y sujetas a validación futura.

Pérdida de interpretabilidad del modelo

Una de las ventajas de las KAN individuales es que permiten inspeccionar las funciones univariantes que modelan cada entrada, ofreciendo una vía directa de interpretabilidad. Sin embargo, al promediar múltiples KANs, esta transparencia se diluye. El modelo resultante pierde la trazabilidad explícita de las transformaciones internas, lo que dificulta la comprensión detallada del comportamiento del sistema, especialmente en aplicaciones donde la explicabilidad es crítica.

7.3 Líneas de investigación futura

A partir de las limitaciones detectadas y los resultados empíricos obtenidos, se proponen diversas líneas de investigación orientadas a potenciar el desempeño, eficiencia e interpretabilidad de RFKAN. Estas propuestas cubren aspectos metodológicos, computacionales y estratégicos, con el objetivo de consolidar el modelo como una herramienta robusta en tareas de regresión tabular.

1. Contraste estadístico completo y validación sobre nuevos conjuntos

Es fundamental completar las diez repeticiones por cada dataset planificado y aplicar la prueba t pareada para calcular valores p que confirmen o refuten con rigor estadístico la mejora observada. Asimismo, ampliar el marco experimental a conjuntos de datos recientes como *California Housing*, *Bike Sharing* o *Credit Default* permitiría reforzar la validez externa del modelo y explorar su comportamiento en dominios contemporáneos y variados.

2. Regularización híbrida y estructuración del modelo

Los hallazgos sugieren que RFKAN podría beneficiarse de una arquitectura híbrida que combine su flexibilidad con un componente lineal explícito. En esta línea,

investigaciones como las de O'Donnell et al. (2020) han mostrado mejoras sustanciales al combinar modelos lineales y redes profundas en esquemas de *stacking*. Además, introducir regularización estructural mediante penalizaciones L2L_2L2 o técnicas como *dropout* —como proponen Gal y Ghahramani (2016) o Zhang et al. (2021)— podría mitigar el riesgo de sobreajuste tanto a nivel de submodelo como del ensamble completo.

3. Optimización computacional y eficiencia de entrenamiento

El coste computacional de entrenar múltiples sub-KANs justifica la exploración de alternativas más eficientes. Entre ellas, destacan la paralelización sobre GPU, la vectorización con bibliotecas como *fun-torch*, o el reemplazo del optimizador LBFGS por AdamW, que ha demostrado un rendimiento competitivo en entornos de aprendizaje profundo.

4. Automatización avanzada de hiperparámetros

La optimización de hiperparámetros mediante rejilla puede sustituirse por métodos más eficientes como *Bayesian Optimization* (Snoek et al., 2012) o *Hyperband* (Levesque et al., 2016). Estas técnicas permiten explorar el espacio de configuraciones de forma más inteligente y eficiente, ajustando la anchura de red, el número de estimadores y los *knots* de manera más precisa.

5. Calibración de la incertidumbre predictiva

La desviación estándar entre submodelos del ensamble RFKAN constituye una fuente natural de estimación de incertidumbre epistemológica. Esta puede calibrarse con *reliability diagrams* y técnicas como *temperature scaling*, tal como proponen Gal y Ghahramani (2016), mejorando la confianza y utilidad de las predicciones, especialmente en contextos de clasificación probabilística.

6. Comparación con métodos de referencia actuales

Para evaluar la competitividad de RFKAN, se recomienda compararlo con métodos de alto rendimiento como *Gradient Boosting Machines*, *XGBoost*, *LightGBM* y *CatBoost*, considerando métricas complementarias como MAE, MAPE o log-loss. Esto permitiría situar al modelo en el panorama actual del aprendizaje tabular con mayor precisión.

7. Interpretabilidad post-hoc

Dado que el ensamble diluye la interpretabilidad directa de las KAN individuales, se sugiere aplicar técnicas como valores SHAP o descomposición ANOVA sobre la salida promedio del modelo. Además, desarrollar visualizaciones que permitan explorar cómo evoluciona la función media del ensamble al añadir submodelos ofrecería una vía intuitiva para comprender su dinámica interna.

7.4 Reflexión Final

La propuesta RFKAN ha demostrado ser una evolución significativa dentro del paradigma de las Kolmogorov–Arnold Networks, al incorporar mecanismos clásicos de ensamblado —como el *bagging* y el muestreo aleatorio de columnas— para mejorar la estabilidad y el rendimiento en tareas de regresión sobre datos tabulares. A través de un protocolo experimental cuidadosamente diseñado y una serie de pruebas empíricas rigurosas, se ha puesto de manifiesto que RFKAN puede reducir el error de predicción respecto a una KAN individual, especialmente en contextos donde la estructura del problema es suficientemente compleja, el tamaño de los datos es moderado o alto, y existe una combinación significativa de variables categóricas y numéricas.

Sin embargo, este modelo no debe entenderse como una solución universal. Los resultados indican claramente que RFKAN puede verse superado por modelos más simples —como

la regresión lineal— en dominios donde predomina la linealidad, o por redes densas como el MLP en escenarios de gran dimensión categórica donde la representación *one-hot* resulta altamente informativa. Del mismo modo, se ha constatado que su desempeño se ve limitado en conjuntos de datos de tamaño reducido, donde el ensamble introduce más varianza de la que es capaz de compensar.

Estas observaciones conducen a una conclusión equilibrada: RFKAN no pretende sustituir a todos los modelos existentes, sino ofrecer una mejora estructural sobre la arquitectura KAN cuando las condiciones del problema lo justifican. Su diseño, que combina la expresividad teórica del teorema de Kolmogórov con herramientas prácticas de la estadística moderna, abre una vía prometedora para el aprendizaje supervisado en tabulares mixtos.

El trabajo realizado aporta una prueba de concepto robusta, un marco experimental reproducible y una base para comparaciones futuras. Aún así, quedan abiertas múltiples vías de mejora: desde la regularización del conjunto hasta su aplicación a tareas más complejas, pasando por la optimización computacional y la incorporación de medidas de incertidumbre bien calibradas. En este sentido, RFKAN no es solo una contribución concreta, sino también una invitación a seguir explorando las sinergias entre métodos clásicos y nuevas arquitecturas funcionales.

Con todo ello, este proyecto culmina consolidando un aporte técnico relevante y ofreciendo un punto de partida claro para futuras investigaciones que aspiren a desarrollar modelos predictivos más robustos, interpretables y ajustados a las características específicas de los datos tabulares reales.

Bibliografía

- Alter, T., Lapid, R., & Sipper, M. (2024). *On the robustness of Kolmogorov-Arnold Networks: An adversarial perspective*. arXiv. <https://arxiv.org/abs/2408.13809>
- Bresson, R., Nikolentzos, G., Panagopoulos, G., Chatzianastasis, M., Pang, J., & Vazirgiannis, M. (2025, March 6). *KAGNNs: Kolmogorov-Arnold Networks meet Graph Learning*. OpenReview. <https://openreview.net/forum?id=03UB1MCAMr>
- Fernández, R., & Ríos, L. (2025). KANs in retail demand forecasting: Explainability and robustness in predictive sales models. *Journal of Business Analytics*, 12(1), 34–49.
- Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. arXiv. <https://arxiv.org/abs/1506.02142>
- Genet, R., & Inzirillo, H. (2024, May 12). *TKAN: Temporal Kolmogorov-Arnold Networks*. arXiv. <https://arxiv.org/abs/2405.07344>
- Kiamari, M., Kiamari, M., & Krishnamachari, B. (2024, June 10). *GKAN: Graph Kolmogorov-Arnold Networks*. arXiv. <https://arxiv.org/abs/2406.06470>
- Levesque, J.-C., Gagné, C., & Sabourin, R. (2016). Bayesian hyperparameter optimization for ensemble learning. arXiv. <https://arxiv.org/abs/1611.07409>
- Liu, Z., Ma, P., Wang, Y., Matusik, W., & Tegmark, M. (2024, August 19). *KAN 2.0: Kolmogorov-Arnold Networks meet science*. arXiv. <https://arxiv.org/abs/2408.10205>
- Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljagic, M., Hou, T. Y., & Tegmark, M. (2025, February). *KAN: Kolmogorov-Arnold Networks*. OpenReview. <https://openreview.net/forum?id=Ozo7qj5vZi>
- Mendoza, F., Hidalgo, M., & Torres, J. (2024). Explainable Kolmogorov-Arnold models in credit scoring under EU regulatory constraints. *AI in Finance*, 3(2), 78–92.
- O'Donnell, A., MacDonald, T., & Rivera, J. (2020). Feature weighted linear stacking through neural networks. *Neurocomputing*, 389, 89–102. <https://doi.org/10.1016/j.neucom.2020.01.112>
- Ortiz-Luna, D., García, P., & Sánchez, H. (2025). Applying Kolmogorov-Arnold Networks in smart manufacturing: Anomaly detection with interpretability. *IEEE Transactions on Industrial Informatics*, 21(3), 2154–2162.
- Shen, H., Zeng, C., Wang, J., & Wang, Q. (2024, July 20). *Reduced effectiveness of Kolmogorov-Arnold Networks on functions with noise*. arXiv. <https://arxiv.org/abs/2407.14882>

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 25, 2951–2959.

Vaca-Rubio, C. J., Blanco, L., Pereira, R., & Caus, M. (2024). *Kolmogorov-Arnold Networks (KANs) for time series analysis*. Semantic Scholar.
<https://www.semanticscholar.org/paper/081eb8781725e560f597b01c624fe65618c3c0f8>

Zhang, X., Wang, M., & Li, J. (2021). A review on dropout regularization approaches for deep neural networks. *Electronics*, 12(14), 3106.
<https://doi.org/10.3390/electronics12143106>

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Diego Pérez-Rasilla Martino, estudiante de E6 Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "Randomized Ensemble of Kolmogorov Arnold Networks" 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. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
4. **Interpretador de código:** Para realizar análisis de datos preliminares.
5. **Estudios multidisciplinares:** Para comprender perspectivas de otras comunidades sobre temas de naturaleza multidisciplinar.
6. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
7. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
8. **Generador de datos sintéticos de prueba:** Para la creación de conjuntos de datos ficticios.
9. **Generador de problemas de ejemplo:** Para ilustrar conceptos y técnicas.
10. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.
11. **Generador de encuestas:** Para diseñar cuestionarios preliminares.
12. **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: 16/06/2025

Firma: _____

