Selección prescriptiva y sostenible de modelos para predicción de series temporales: ¿cuándo es necesario usar Deep Learning?

Pablo Reina-Jiménez

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Sevilla,
Av. Reina Mercedes S/N, 41012 Sevilla, España
prjimenez@us.es

Directores: Dr. José C. Riquelme y Dra. María Martínez-Ballesteros
Fecha de inicio de la tesis: Octubre de 2024

Resumen La predicción de series temporales es una tarea central en numerosos dominios, desde la gestión energética y el tráfico hasta la monitorización ambiental, la economía o la industria. En los últimos años, los modelos de *Deep Learning* han alcanzado resultados destacados gracias a su capacidad para aprender representaciones complejas y dependencias temporales de largo alcance. Sin embargo, esta mejora suele ir acompañada de un aumento considerable del coste computacional, del consumo energético, de las emisiones asociadas y de una menor interpretabilidad. Esta tesis doctoral parte de la hipótesis de que no todos los escenarios de predicción justifican el uso de modelos profundos: existen conjuntos de datos y configuraciones temporales en los que modelos más simples, especialmente métodos basados en árboles, pueden alcanzar una precisión estadísticamente comparable con un coste muy inferior.
El objetivo principal de la investigación es caracterizar los escenarios en los que resulta necesario utilizar modelos de *Deep Learning* para predicción de series temporales y aquellos en los que alternativas más eficientes son suficientes. Para ello se propone una metodología basada en benchmarking a gran escala, análisis estadístico bayesiano, evaluación de eficiencia computacional y sostenibilidad, extracción de metacaracterísticas de las series temporales y construcción de un marco prescriptivo e interpretable de selección de modelos. La aportación original esperada consiste en transformar la selección de modelos de predicción, tradicionalmente guiada por intuición o por comparativas aisladas, en un proceso automático, explicable y sensible al equilibrio entre precisión, eficiencia y sostenibilidad.

Keywords: Series temporales · Selección de modelos · Deep Learning · Modelos basados en árboles · Green AI · Meta-learning · Inteligencia Artificial interpretable

1. Introducción y motivación

La predicción de series temporales constituye un problema fundamental en inteligencia artificial y análisis de datos. Su relevancia se observa en dominios

como la planificación energética, la predicción de demanda, la movilidad, la meteorología, la economía, la salud pública o la monitorización ambiental. En todos estos casos, disponer de predicciones precisas permite anticipar recursos, optimizar decisiones y reducir riesgos operativos.

La investigación ha evolucionado desde métodos estadísticos clásicos hacia técnicas de *Machine Learning* y arquitecturas de *Deep Learning*. Modelos como LSTM, CNN, N-BEATS, Transformer o iTransformer capturan patrones no lineales, dependencias de largo alcance e interacciones multivariantes, pero suelen requerir más datos, más cómputo, búsqueda de hiperparámetros y menor transparencia que alternativas más simples.

Esta situación genera la pregunta central de la tesis: *¿cuándo merece realmente la pena utilizar un modelo de Deep Learning?* Una mejora marginal en error puede no compensar si multiplica el coste computacional, el consumo energético o las emisiones. En escenarios reales también importan el reentrenamiento, el despliegue, la interpretabilidad y la disponibilidad de recursos.

La motivación de esta tesis es avanzar hacia una selección de modelos prescriptiva y sostenible: no solo comparar algoritmos, sino identificar qué características de la serie o del problema justifican modelos profundos y cuáles permiten usar modelos más ligeros sin pérdida significativa de precisión.

2. Antecedentes e investigaciones previas

La predicción de series temporales ha evolucionado desde modelos estadísticos clásicos hacia métodos de *Machine Learning*, arquitecturas de *Deep Learning* y, más recientemente, modelos fundacionales. Los enfoques estadísticos, como ARIMA o el suavizado exponencial, han sido durante décadas una referencia por su interpretabilidad, bajo coste y sólida formulación matemática [2]. Sin embargo, suelen depender de supuestos como linealidad, estacionariedad o patrones temporales relativamente estables, lo que limita su capacidad ante series con no linealidades, múltiples variables, cambios de régimen o dependencias complejas.

Los modelos basados en árboles son una alternativa eficiente dentro del aprendizaje automático clásico. Árboles de regresión, M5P, Random Forest y XGBoost permiten capturar relaciones no lineales con menor coste que muchas arquitecturas profundas [3,15,5]. En series temporales suelen aplicarse transformando la serie en un problema supervisado mediante retardos, ventanas o variables derivadas. Esta estrategia es competitiva, aunque puede ser limitada para dependencias largas si no se codifican explícitamente.

El *Deep Learning* permitió aprender representaciones directamente desde los datos. LSTM, CNN, N-BEATS y Transformers han ampliado la capacidad para modelar dependencias temporales, patrones locales, estructuras multivariantes y horizontes largos [8,9,14,17,10]. Sin embargo, las revisiones recientes señalan que su rendimiento depende del dominio, del horizonte, del tamaño de los datos y de la configuración experimental [4].

Competiciones y benchmarks como M4 y M5 han comparado métodos estadísticos, aprendizaje automático y modelos profundos sobre grandes colecciones de series [11,12]. Sus resultados sugieren que no existe una familia dominante en todos los escenarios. Además, muchas comparativas se centran en dominios concretos, por lo que sus conclusiones no siempre son transferibles cuando cambian la frecuencia, longitud, dimensionalidad, ruido, estacionalidad u horizonte de predicción.

Muchas comparativas priorizan el error y dejan en segundo plano entrenamiento, inferencia, energía, emisiones e interpretabilidad. La perspectiva de *Green AI* propone medir la calidad de un modelo también por el coste necesario para alcanzarla [16]. En series temporales, con reentrenamientos frecuentes y despliegue continuo, reportar consumo energético y huella de carbono resulta especialmente relevante [7].

El *meta-learning* permite automatizar la selección de modelos caracterizando las series mediante metacaracterísticas y aprendiendo qué método es más adecuado para cada caso, como en FFORMA [13]. Esta idea conecta con el teorema *No Free Lunch*, según el cual ningún algoritmo es óptimo para todos los problemas [18]. Sin embargo, estas aproximaciones no siempre integran eficiencia, sostenibilidad e interpretabilidad en una recomendación prescriptiva.

Finalmente, los modelos fundacionales para series temporales plantean el preentrenamiento sobre grandes colecciones para favorecer transferencia y predicción *zero-shot* o con poco ajuste específico [1,6]. Esta línea es prometedora para nuevos dominios y horizontes largos, pero refuerza la pregunta central: cuándo su complejidad está justificada frente a modelos más simples, eficientes e interpretables.

En conjunto, el estado del arte muestra numerosas familias de modelos y comparativas empíricas, pero no un marco unificado que relacione las características de las series temporales con la conveniencia de utilizar modelos simples, profundos o fundacionales bajo criterios conjuntos de precisión, eficiencia, sostenibilidad e interpretabilidad. Esta ausencia de guía prescriptiva constituye el punto de partida de la tesis.

3. Hipótesis y objetivos

3.1. Hipótesis de partida

La hipótesis principal de esta tesis es que la necesidad de utilizar modelos de *Deep Learning* en predicción de series temporales depende de las características del problema, y que en un número significativo de escenarios los modelos basados en árboles pueden ofrecer una precisión estadísticamente comparable con un coste computacional y energético muy inferior. Esta hipótesis no implica descartar los modelos profundos, sino determinar de forma sistemática cuándo su mayor capacidad de representación está justificada y cuándo una alternativa más simple resulta suficiente.

Esta hipótesis se desglosa en las siguientes hipótesis específicas:

- **H1.** No existe una familia de modelos que domine de forma universal en todos los escenarios de predicción de series temporales. La conveniencia de un modelo depende del dominio, la frecuencia de muestreo, la longitud de la serie, el número de variables, la presencia de estacionalidad, el nivel de ruido y el horizonte de predicción.
- **H2.** Las diferencias de precisión entre modelos profundos y modelos basados en árboles son, en muchos casos, menores que las diferencias de coste computacional, tiempo de entrenamiento y consumo energético.
- **H3.** Las características intrínsecas de las series temporales, como complejidad, tendencia, estacionalidad, ruido, dimensionalidad, autocorrelación, longitud disponible e irregularidad, permiten anticipar qué familia de modelos será más adecuada.
- **H4.** A medida que aumentan el horizonte de predicción, la dimensionalidad y la complejidad temporal, pueden aparecer escenarios en los que modelos profundos o fundacionales estén justificados, pero esta ventaja debe evaluarse frente a modelos más simples bajo criterios de precisión, eficiencia, sostenibilidad e interpretabilidad.
- **H5.** Es posible construir un marco prescriptivo interpretable que recomiende modelos de predicción considerando simultáneamente las características de los datos, el horizonte de predicción, el coste computacional, el consumo energético, las emisiones asociadas y la precisión esperada.

3.2. Objetivo principal

El objetivo principal de la tesis es desarrollar y validar un marco prescriptivo, interpretable y sostenible para seleccionar modelos de predicción de series temporales, identificando cuándo está justificado utilizar modelos de *Deep Learning* o modelos fundacionales y cuándo son preferibles modelos más simples, como los basados en árboles, por ofrecer un equilibrio más favorable entre precisión, coste y sostenibilidad.

3.3. Subobjetivos

Para alcanzar el objetivo principal, se plantean los siguientes subobjetivos:

1. Revisar y organizar la literatura sobre predicción de series temporales, selección de modelos, *Green AI*, meta-learning, interpretabilidad y modelos fundacionales, identificando la ausencia de marcos unificados de comparación.
2. Diseñar un protocolo experimental reproducible para comparar modelos basados en árboles, modelos de *Deep Learning* y, en fases posteriores, modelos fundacionales, en escenarios univariantes y multivariantes.
3. Construir y ampliar progresivamente un conjunto de problemas de predicción que incluya distintos dominios, frecuencias, longitudes, niveles de dimensionalidad y horizontes, evaluando modelos basados en árboles, arquitecturas profundas y modelos fundacionales.

4. Analizar el rendimiento de los modelos no solo en términos de error predictivo, sino también de tiempo de entrenamiento, tiempo de inferencia, consumo energético, emisiones de CO_2 y facilidad de interpretación.
5. Aplicar tests estadísticos bayesianos y regiones de equivalencia práctica para distinguir mejoras relevantes de diferencias numéricas pequeñas que no justifiquen un aumento de complejidad.
6. Extraer metacaracterísticas de las series temporales y construir un metamodelo interpretable que relacione dichas características con la familia de modelos más adecuada.
7. Generar reglas de decisión comprensibles y validar la utilidad del marco en nuevos datasets, dominios y horizontes largos, analizando su generalización y transferencia.

4. Metodología y plan de trabajo

La metodología de la tesis se plantea como una secuencia incremental: formalización del problema, diseño experimental común, comparación inicial de modelos, ampliación a un benchmark de mayor escala, extensión a nuevos dominios y modelos fundacionales, e integración final en un marco prescriptivo interpretable.

4.1. Bloque 1: Revisión del estado del arte y formalización del problema

El primer bloque consiste en estudiar la evolución de los métodos de predicción de series temporales, desde modelos estadísticos clásicos hasta modelos de *Machine Learning*, *Deep Learning* y modelos fundacionales. Esta revisión no se limitará a la precisión predictiva, sino que organizará la literatura en torno a cinco dimensiones: exactitud, eficiencia computacional, interpretabilidad, sostenibilidad y capacidad de generalización entre dominios.

El resultado esperado es una taxonomía de métodos, escenarios y criterios de evaluación que sitúe la tesis frente a la falta de un marco unificado para decidir cuándo usar modelos simples, profundos o fundacionales.

4.2. Bloque 2: Diseño experimental y construcción progresiva de datasets

El segundo bloque consiste en definir un protocolo experimental homogéneo para evaluar los modelos en diferentes tipos de series temporales. El protocolo contemplará:

- selección de datasets reales procedentes de distintos dominios;
- construcción de ventanas temporales mediante históricos pasados y horizontes de predicción;
- normalización de los datos y partición en entrenamiento y prueba;

- definición de configuraciones univariantes y multivariantes;
- variación de horizontes de predicción y tamaños de histórico;
- búsqueda sistemática de hiperparámetros;
- registro homogéneo de errores, tiempos, consumo energético y emisiones estimadas.

El bloque evolucionará desde datasets multivariantes iniciales hacia un benchmark de mayor escala y, finalmente, hacia nuevas colecciones, dominios y horizontes largos para estudiar la generalización del marco.

4.3. Bloque 3: Benchmarking de modelos

El tercer bloque se centra en la ejecución de experimentos comparativos bajo un mismo protocolo. La tesis no se limitará a un único benchmark, sino que desarrollará una comparación por fases:

- **Fase inicial:** comparación entre M5P, LSTM y CNN en series multivariantes, con el objetivo de analizar si las mejoras de precisión de los modelos profundos compensan su mayor tiempo de entrenamiento.
- **Fase de ampliación:** incorporación de Regression Tree, Random Forest, XGBoost, N-BEATS, Transformer e iTransformer, junto con un conjunto más amplio de series temporales y problemas de predicción.
- **Fase futura:** evaluación de modelos fundacionales para series temporales en nuevas series, nuevos dominios y horizontes de predicción más largos, comparándolos con modelos basados en árboles y modelos profundos entrenados específicamente para cada problema.

La evaluación considerará precisión predictiva (MAE, MAPE, RMSE y WAPE), eficiencia computacional (entrenamiento e inferencia), sostenibilidad (energía y CO_2), robustez estadística (tests bayesianos y regiones de equivalencia práctica) y utilidad práctica (interpretabilidad, despliegue y adaptación a nuevos dominios). Así, se cuantificará no solo qué modelo obtiene menor error, sino también cuánto cuesta obtener esa mejora.

4.4. Bloque 4: Extracción de metacaracterísticas y meta-learning

El cuarto bloque busca explicar los resultados obtenidos en los benchmarks. Para ello, cada serie temporal o problema de predicción se describirá mediante un conjunto de metacaracterísticas. Estas incluirán, entre otras, medidas de media, varianza, asimetría, curtosis, autocorrelación, tendencia, estacionalidad, estacionariedad, entropía, irregularidad, longitud de la serie, dimensionalidad y horizonte de predicción.

Estas metacaracterísticas se utilizarán como entrada de un meta-modelo cuyo objetivo será recomendar el algoritmo más adecuado para cada escenario. Se priorizarán modelos interpretables, especialmente árboles de decisión, porque permiten obtener reglas explícitas que indiquen cuándo basta un modelo basado en árboles y cuándo puede estar justificado utilizar un modelo profundo o fundacional. La finalidad no es únicamente automatizar la selección, sino también explicar la recomendación.

4.5. Bloque 5: Validación, análisis de relevancia y transferencia

El último bloque se dedicará a validar el marco propuesto, analizar sus limitaciones y estudiar su aplicabilidad en dominios reales. Se evaluará si las reglas aprendidas generalizan a nuevos datasets y si las recomendaciones permiten reducir coste computacional sin pérdida significativa de precisión.

La validación incluirá series no vistas, horizontes más largos y escenarios donde los modelos fundacionales puedan aportar transferencia entre dominios, evaluando también cuándo su coste no está justificado.

También se estudiará la utilidad del sistema para usuarios no expertos, que podrían emplear el marco como apoyo para seleccionar modelos sin necesidad de ejecutar comparativas exhaustivas.

4.6. Plan temporal e hitos

El plan temporal considera una tesis iniciada en octubre de 2024 y se presenta como una planificación completa del trabajo doctoral. Los hitos avanzan desde la validación inicial del trade-off precisión/eficiencia hasta el benchmark ampliado, la incorporación de modelos fundacionales, la consolidación del marco prescriptivo y la redacción de la memoria.

Oct. 2024 – Jun. 2025.
Revisión bibliográfica inicial sobre predicción de series temporales, modelos basados en árboles, *Deep Learning*, eficiencia y *Green AI*. Definición del primer protocolo experimental para series multivariantes. Comparación inicial entre M5P, LSTM y CNN. Aplicación de tests bayesianos para analizar error y tiempo de cómputo. **Hito:** primer estudio sobre el trade-off entre precisión y eficiencia, previsto como publicación en IWANN.

Jul. 2025 – Ene. 2026.
Ampliación del benchmark a un conjunto mucho mayor de problemas y series temporales. Incorporación de nuevos modelos, incluyendo Regression Tree, Random Forest, XGBoost, N-BEATS, Transformer e iTransformer. Inclusión de medidas de consumo energético y emisiones de CO_2. Extracción sistemática de metacaracterísticas de las series y construcción inicial del marco prescriptivo. **Hito:** segundo estudio centrado en selección de modelos basada en precisión, eficiencia, sostenibilidad e interpretabilidad, previsto como artículo en ASOC.

Feb. 2026 – Feb. 2027.
Extensión del estudio hacia nuevas colecciones de series temporales y nuevos dominios de aplicación. Rediseño del protocolo para considerar horizontes de predicción más largos, distintos tamaños de histórico y escenarios en los que la transferencia entre dominios pueda ser relevante. Incorporación y evaluación de modelos fundacionales para series temporales, comparándolos con modelos

basados en árboles, modelos profundos entrenados desde cero y el marco prescriptivo ya desarrollado. **Hito:** benchmark ampliado para escenarios de largo horizonte y modelos fundacionales.

Mar. 2027 – Ene. 2028.
Integración de los resultados anteriores en un marco de recomendación más general, capaz de considerar modelos basados en árboles, arquitecturas profundas y modelos fundacionales. Validación final de reglas interpretables por dominio, tipo de serie y horizonte de predicción. Análisis de limitaciones, preparación de material reproducible, consolidación de publicaciones y redacción de la memoria. **Hito:** depósito y defensa de la tesis doctoral.

Además de estos hitos, se prevén tareas transversales durante toda la tesis: mantenimiento del código experimental, documentación de datasets y configuraciones, revisión continua de literatura, preparación de artículos derivados y participación en congresos. La planificación se mantendrá flexible para incorporar nuevos modelos fundacionales o benchmarks que contribuyan a responder cuándo merece la pena aumentar la complejidad del modelo.

5. Resultados preliminares

Los resultados preliminares apoyan la hipótesis de partida. En el primer estudio, LSTM y CNN obtuvieron errores algo menores en algunos datasets, especialmente en series complejas, pero las diferencias de precisión no siempre fueron significativas y los tiempos de entrenamiento sí aumentaron de forma clara. M5P mostró una precisión competitiva y una eficiencia superior, incluso sin GPU.

Además, el análisis estadístico mediante tests bayesianos permitió distinguir entre diferencias reales y diferencias prácticamente equivalentes. Este punto es fundamental porque evita concluir que un modelo es mejor solo por obtener una mejora numérica pequeña en una métrica de error. En el contexto de la tesis, una mejora marginal solo debe considerarse relevante si compensa el coste adicional en tiempo, energía y complejidad.

El segundo estudio amplía el análisis a más de 50.000 series y 31 problemas. Los resultados muestran que la superioridad de los modelos profundos no es universal: en numerosos escenarios, modelos como M5P o Regression Tree alcanzan resultados estadísticamente comparables con tiempos y consumos energéticos muy inferiores. Las arquitecturas profundas pueden requerir uno o dos órdenes de magnitud más energía sin una mejora proporcional en precisión.

Estos resultados preliminares refuerzan la necesidad de un marco de selección que tenga en cuenta simultáneamente precisión, coste computacional y sostenibilidad. La contribución de la tesis consiste en utilizar estos resultados no solo para comparar modelos, sino para aprender reglas prescriptivas que orienten la selección futura.

6. Relevancia científica y transferencia

6.1. Relevancia científica

La relevancia científica de esta tesis se sitúa en tres niveles. En primer lugar, aborda una cuestión metodológica abierta en predicción de series temporales: la ausencia de criterios sistemáticos para decidir cuándo utilizar modelos complejos. Aunque existen numerosos benchmarks, muchos se centran en comparar errores promedio y no explican bajo qué condiciones cada modelo es recomendable.

En segundo lugar, la tesis incorpora una perspectiva de sostenibilidad al proceso de selección de modelos. Esta perspectiva es especialmente relevante en un contexto en el que la inteligencia artificial avanza hacia arquitecturas cada vez más grandes y costosas. La tesis defiende que la eficiencia no debe considerarse un aspecto secundario, sino una dimensión central de la calidad de un modelo.

En tercer lugar, la propuesta combina benchmarking, análisis estadístico, metacaracterísticas y modelos interpretables de meta-learning. Esta combinación permite pasar de una evaluación descriptiva a una herramienta prescriptiva. En lugar de limitarse a afirmar qué modelo fue mejor en un conjunto de datos concreto, la tesis busca explicar por qué lo fue y cuándo cabe esperar que vuelva a serlo.

6.2. Aplicación y transferencia a la sociedad

Desde el punto de vista aplicado, el marco propuesto puede ser útil en cualquier entorno donde se realicen predicciones periódicas sobre series temporales. Algunos ejemplos son:

- empresas energéticas que predicen demanda o producción renovable;
- administraciones públicas que monitorizan tráfico, contaminación o consumo;
- sistemas industriales que necesitan mantenimiento predictivo;
- plataformas de comercio que estiman ventas o demanda;
- organizaciones con recursos computacionales limitados que necesitan modelos eficientes y desplegables.

El impacto social se deriva principalmente de la reducción de costes y de la promoción de una inteligencia artificial más sostenible. En muchos casos, utilizar modelos más simples puede permitir que organizaciones con menos recursos accedan a sistemas predictivos eficaces sin necesidad de infraestructura especializada. Además, reducir entrenamientos innecesarios de modelos profundos contribuye a disminuir el consumo energético asociado a la inteligencia artificial.

6.3. Aportación original de la tesis

La aportación original de la tesis no se limita a comparar modelos, sino que propone un cambio de enfoque: pasar de la pregunta *¿qué modelo tiene menor error?* a la pregunta *¿qué modelo conviene utilizar en este escenario concreto considerando precisión, coste, sostenibilidad e interpretabilidad?*

Esta aportación se concreta en:

- un benchmark amplio y reproducible de modelos basados en árboles y modelos de *Deep Learning*;
- una evaluación conjunta de precisión, eficiencia y sostenibilidad;
- una caracterización de escenarios de predicción mediante metacaracterísticas;
- un marco prescriptivo interpretable para recomendar modelos;
- reglas de decisión que puedan ser utilizadas por investigadores y profesionales.

De este modo, la tesis contribuye tanto al avance metodológico de la predicción de series temporales como a la construcción de sistemas de inteligencia artificial más responsables, explicables y sostenibles.

7. Conclusiones esperadas

Se espera que la tesis demuestre que la selección de modelos para series temporales debe abordarse como un problema condicionado por los datos y por los requisitos de uso. Los modelos de *Deep Learning* y los modelos fundacionales son herramientas muy potentes y necesarias en ciertos escenarios, especialmente cuando existen patrones complejos, múltiples variables, grandes volúmenes de datos, horizontes largos y dependencias temporales difíciles de capturar con modelos simples. Sin embargo, no deberían considerarse la opción por defecto cuando modelos más ligeros ofrecen resultados equivalentes con menor coste.

La tesis aspira a proporcionar evidencia experimental y una herramienta metodológica para decidir de forma razonada cuándo utilizar modelos profundos. En lugar de promover una familia de algoritmos concreta, la contribución busca favorecer una selección informada, interpretable y sostenible. Con ello, la investigación se alinea con los principios de la *Green AI* y con la necesidad creciente de desarrollar sistemas predictivos precisos, eficientes y transferibles a contextos reales.

Agradecimientos Esta investigación ha sido financiada por la ayuda PID2023-146037OB-C21, financiada por MICIU/AEI/10.13039/501100011033. Los autores desean agradecer al Ministerio de Universidades de España el contrato predoctoral FPU (FPU23/03390).

Declaración de intereses El autor declara no tener conflictos de intereses relevantes para el contenido de esta propuesta.

Referencias

1. Ansari, A.F., et al.: Chronos: Learning the language of time series (2024). https://doi.org/10.48550/arXiv.2403.07815
2. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. John Wiley & Sons (2015)
3. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
4. Casolaro, A., Capone, V., Iannuzzo, G., Camastra, F.: Deep learning for time series forecasting: Advances and open problems. Information **14**(11), 598 (2023). https://doi.org/10.3390/info14110598

5. Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). pp. 785–794. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939785
6. Das, A., Kong, W., Sen, R., Zhou, Y.: A decoder-only foundation model for time-series forecasting. In: Proceedings of the 41st International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research, vol. 235, pp. 10148–10167. PMLR (2024), https://proceedings.mlr.press/v235/das24c.html
7. Henderson, P., et al.: Towards the systematic reporting of the energy and carbon footprints of machine learning. Journal of Machine Learning Research **21**(248), 1–43 (2020), https://www.jmlr.org/papers/v21/20-312.html
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015). https://doi.org/10.1038/nature14539
10. Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., Long, M.: iTransformer: Inverted transformers are effective for time series forecasting. In: Proceedings of the 12th International Conference on Learning Representations (ICLR) (2024). https://doi.org/10.48550/arXiv.2310.06625, https://openreview.net/forum?id=8QmhwMPmqX
11. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The M4 competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting **36**(1), 54–74 (2020). https://doi.org/10.1016/j.ijforecast.2019.04.014
12. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting **38**(4), 1346–1364 (2022). https://doi.org/10.1016/j.ijforecast.2021.11.013
13. Montero-Manso, P., Athanasopoulos, G., Hyndman, R.J., Talagala, T.S.: Fforma: Feature-based forecast model averaging. International Journal of Forecasting **36**(1), 86–92 (2020). https://doi.org/10.1016/j.ijforecast.2019.02.011
14. Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: Proceedings of the 8th International Conference on Learning Representations (ICLR) (2020), https://openreview.net/forum?id=r1ecqn4YwB
15. Quinlan, J.R.: Learning with continuous classes. In: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (AI 1992). pp. 343–348. World Scientific (1992), https://www.worldscientific.com/doi/pdf/10.1142/1897#page=356
16. Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O.: Green AI. Communications of the ACM **63**(12), 54–63 (2020). https://doi.org/10.1145/3381831
17. Vaswani, A., et al.: Attention is all you need. In: Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS). vol. 30 (2017). https://doi.org/10.5555/3295222.3295349
18. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE transactions on evolutionary computation **1**(1), 67–82 (2002)