



La aplicación de datos masivos en economía de la energía: una revisión

MIGUEL ÁNGEL RODRÍGUEZ LÓPEZ
DIEGO RODRÍGUEZ RODRÍGUEZ

Documento de Trabajo 2024/08
Agosto de 2024

fedea

Las opiniones recogidas en este documento son las de sus autores y no coinciden necesariamente con las de Fedea.

La aplicación de datos masivos en economía de la energía: una revisión¹

Miguel Ángel Rodríguez López (ACCIONA)

Diego Rodríguez Rodríguez (UCM y FEDEA)

(versión 5 de julio de 2024)

1. Introducción

La energía es un recurso básico para la realización de cualquier actividad y, como tal, ha sido tradicionalmente objeto del análisis económico. Sin embargo, la relevancia de este campo de estudio se acrecienta al considerar los retos a los que se enfrenta el conjunto del planeta para limitar el proceso de cambio climático y mitigar sus consecuencias. En ello la energía juega un papel fundamental, ya que en torno al 80% de las emisiones de gases de efecto invernadero (GEI) están asociadas con la generación y consumo de la energía (IEA, 2023a). Este porcentaje incluye la generación eléctrica que hace uso de hidrocarburos, pero sobre todo el uso de estos en el transporte, en las actividades industriales y de servicios, así como por parte de los hogares. La transición energética, entendida como el cambio en los procesos de generación y usos de la energía, se enfrenta con formidables retos tecnológicos y de mercado, cuyos efectos sobre los agentes y sobre el territorio probablemente no serán homogéneos. Adicionalmente, en los últimos años se han puesto de manifiesto los problemas asociados a los riesgos en el suministro energético, con tensionamientos en los mercados por los aumentos de los precios y consecuencias sobre las rentas y las decisiones de los hogares y de las empresas. En este contexto crecientemente complejo, la modelización económica de la energía contribuye a generar conocimientos y herramientas que permitan una gestión más eficiente de los sistemas energéticos y, en suma, ayuden a avanzar en el proceso de descarbonización. A ese respecto, el hecho de que en la actualidad se definan objetivos y actuaciones de “energía y clima” es un indicador claro de la íntima asociación existente entre ambas cuestiones.

Este proceso de transición energética se desarrolla de modo paralelo a la digitalización del conjunto de actividades económicas. De hecho, la existencia de sinergias entre la transición energética y la transición digital ha conducido a acuñar el término de “transición gemela” (Bianchini et al., 2023; Huang et al., 2024). La digitalización proporciona un conjunto de datos mucho más amplio que el previamente disponible como consecuencia de la existencia de múltiples puntos de medición situados a lo largo de las cadenas de producción, distribución y consumo de energía. Ello ha facilitado la aplicación en este campo de las diversas técnicas de análisis masivos de datos (Zhou et al., 2016).

La aplicación práctica de enfoques “conducidos por los datos” (*big data-driven*) en la gestión energética abarca áreas muy diversas, con soluciones que están transformando radicalmente la forma en que se planifica, opera y optimiza las infraestructuras energéticas. Por ejemplo, el diseño de estrategias inteligentes para maximizar la producción de energía renovable, la predicción más precisa de la demanda o la integración efectiva de las energías renovables en la red eléctrica, todos ellos elementos clave para asegurar la estabilidad y el funcionamiento más eficiente del sistema energético. Asimismo, el análisis de datos masivos por tipos de energía, zonas geográficas y tipologías de consumidores permite la formulación de políticas y decisiones estratégicas en la planificación urbana y el transporte sostenible al revelar patrones de movilidad (Abuljadail et al., 2023; Lonergan et al., 2023), proporciona información

¹ Este trabajo forma parte del libro *Nuevos métodos de investigación económica con datos masivos* (Daniel Peña, ediot), de próxima publicación.

valiosa sobre los hábitos de consumo energético o facilita el diseño de políticas dirigidas a impulsar la transición hacia infraestructuras más sostenible (Huntington, 2021). También permite a las empresas tomar decisiones más informadas para sus decisiones de inversión asociadas con la transición energética, como el refuerzo de la red eléctrica y la identificación de las mejores ubicaciones para el despliegue de los puntos de recarga para vehículos eléctricos, o facilita la implementación por parte de los reguladores de incentivos para la adopción de tecnologías más limpias, entre otros muchos aspectos. En el mismo sentido, la simulación de escenarios energéticos basados en datos masivos es una herramienta poderosa para evaluar el impacto de políticas y decisiones estratégicas a largo plazo (Rodríguez-López, 2022). Al modelar diferentes configuraciones de redes energéticas y escenarios de adopción de tecnologías, se puede avanzar en la identificación de las mejores estrategias para lograr una transición energética más eficiente en costes y más sostenible (Taniguchi-Matsuoka et al., 2020; Ba et al., 2023).

Asimismo, el hecho de poder predecir con menor error eventos futuros es especialmente relevante ya que permite reducir los costes de la transición. Un claro ejemplo es el mantenimiento predictivo y prescriptivo de las instalaciones, que ha sido tradicionalmente una pieza clave en la gestión de activos en el sector energético. En ese ámbito, la aplicación de técnicas de aprendizaje automático permite predecir posibles fallos en los equipos de generación y distribución de energía. Ello no solo facilita reducir el riesgo de interrupciones no planificadas sino que, también, posibilita optimizar los recursos al programar el mantenimiento en el momento preciso. Un caso de uso habitual es la implementación de sistemas de monitoreo en plantas de generación solares o eólicas en las que, a través de la utilización de datos históricos y algoritmos de Inteligencia Artificial (IA), se puede predecir el rendimiento futuro y anticipar posibles problemas, permitiendo una acción proactiva para resolverlos (Rodríguez-López, 2015). La utilización de algoritmos avanzados permite también implementar estrategias para optimizar el uso de la energía y, en ese sentido, una gestión más inteligente de la demanda energética. Ello incluye, por ejemplo, programas de incentivos para los consumidores que reduzcan su demanda en momentos de alta congestión en la red eléctrica, ayudando así a estabilizarla y reduciendo el riesgo de falta de suministro.

El objetivo de este trabajo es ofrecer un panorama general sobre los principales campos de aplicación de las técnicas de datos masivos en el ámbito de la energía. Como podrá comprobarse, esos ámbitos son muy variados, por lo que el trabajo pretende también aportar intuición sobre el tipo de cuestiones que generan más interés o que probablemente recibirán más atención en los próximos años, dada su relevancia en el ámbito de la transición energética. Para ello, en el segundo apartado se realiza una aproximación muy introductoria a la transición energética y al estado de desarrollo de distintas tecnologías y estrategias que son claves en el proceso de descarbonización. A continuación, en el tercer apartado, se describen de un modo sencillo las principales aproximaciones metodológicas para el uso masivo de datos. El apartado cuarto motiva y muestra la aplicación de esas técnicas en diversas áreas claves en el área de la transición energética en el ámbito de la electricidad, como la predicción de la generación y los precios, la agregación de la demanda, la integración de las baterías o la demanda de calefacción, entre otros. El quinto apartado extiende esa aproximación a otros vectores no eléctricos, en particular el gas y el petróleo. Por último, el sexto apartado cierra el trabajo con algunas reflexiones finales.

2. La transición energética y el uso de datos masivos en energía

No hay duda alguna en la comunidad científica acerca de que es imprescindible una transformación profunda del escenario energético actual como herramienta básica para abordar la crisis climática. La evidencia contenida en los informes elaborados por los expertos del Panel Intergubernamental de Cambio Climático alerta con rotundidad sobre la necesidad de proceder rápidamente a la descarbonización de la

oferta y la demanda energética. Con ese fin, los países se están comprometiendo a alcanzar objetivos ambiciosos en la descarbonización de su mix energético, que implican el desplazamiento de los combustibles fósiles dentro de la matriz de consumo energético primario y la adopción de fuentes de energía renovables y otros vectores como el hidrógeno. Ello se acompañaría con un proceso de electrificación del consumo final de energía, con reducciones de la intensidad energética (consumo por unidad de valor generado) y, tal vez, con la incorporación de tecnologías de captura y reutilización del carbono.

El proceso de integración de fuentes renovables, que son intermitentes por su propia naturaleza, implica a su vez un nuevo papel para la gestión de la demanda y el almacenamiento de energía. Adicionalmente, la electrificación de sectores como el transporte y la industria son componentes esenciales para reducir la dependencia de los combustibles fósiles. Sin embargo, en el estado actual y previsible de la tecnología para los próximos años, la descarbonización requerirá también del impulso de los gases renovables, como el hidrógeno y biogás/biometano. Aunque la reducción de emisiones netas vinculada a los biogases sea inferior a la que ofrecería una solución electrificada, lo cierto es que hay importantes barreras tecnológicas para la descarbonización de algunos sectores de actividad, como el transporte marítimo o el aéreo. En este último caso, por ejemplo, que es responsable de un porcentaje pequeño (en torno al 2,5%) pero creciente de las emisiones mundiales, la utilización de combustibles “sostenibles” es la única vía realista para avanzar en la reducción de emisiones netas. Paralelamente, se buscan medidas para reducir la intensidad energética mediante la implementación de tecnologías más eficientes y prácticas de gestión energética inteligente.

Como se analizará en este trabajo, todas estas actuaciones requieren una gestión más inteligente del sistema energético en su conjunto, lo que resulta inviable sin el uso del dato y sin modelos matemáticos que permitan adelantarse en la medida de lo posible a los acontecimientos (Entezari et al., 2023). Sin embargo, aunque el panorama general de posibles estrategias es claro, los resultados que se van alcanzando en cada uno de los distintos componentes de esa estrategia general, así como los resultados globales alcanzados por los países, se alejan en menor o mayor medida de las sendas de penetración y de descarbonización planteadas (IEA, 2023b).

En ese contexto, el informe de IRENA (2023a) ofrece un panorama actualizado sobre cuál es la situación actual en relación con las inversiones y los objetivos planteados en los horizontes 2030 y 2050 para lograr el objetivo último de alcanzar un incremento de temperaturas por debajo de 1,5 °C con respecto al periodo preindustrial. El informe muestra que hay algunos componentes de la transición que siguen las sendas de penetración deseadas, pero también que hay otros muchos que se sitúan claramente por debajo de las sendas requeridas. Así, en relación con la participación de las renovables en el mix de generación eléctrica, el informe es moderadamente optimista acerca de que pueda alcanzarse un porcentaje del 68% para 2030 y del 91% para 2050. Ese moderado optimismo se fundamenta en el importante despliegue global de nueva capacidad de generación renovable, que en cualquier caso debería acelerarse para cumplir con los objetivos. Es importante señalar que la electricidad representa actualmente en torno el 22%, del consumo final de energía, con metas del 29% y 51% para 2030 y 2050 en el escenario objetivo. La utilización directa de renovables para usos finales y producción de calor, que actualmente representa un 17% de esos consumos, debería alcanzar el 35% para 2030 y el 82% para 2050.

En el sector del transporte, la electrificación también avanza con dificultades. Los escenarios de IRENA que son compatibles con un incremento por debajo de 1,5° requerirían pasar del parque actual de 10,5 millones de coches eléctricos en circulación a un parque de 360 millones para 2030 y 2.180 millones para 2050. Ello requiere de un avance mucho más rápido del despliegue de puntos de carga. Sin embargo, la

visión de IRENA es algo más optimista en el ámbito de las inversiones requeridas para el despliegue de las bombas de calor, que son clave para la descarbonización del consumo energético de los hogares. Asimismo, el despliegue del hidrógeno renovable, que actualmente se encuentra en una fase muy inicial, tendría que acelerarse extraordinariamente para pasar desde los 0,7 millones de toneladas/año de la actualidad a los niveles del escenario objetivo: 125 y 523 millones de toneladas/año en 2030 y 2050, respectivamente. Otro caso donde hay una distancia enorme entre el nivel actual y las necesidades proyectadas se refiere a la captura y almacenamiento de carbono (CCS), aún en fase de desarrollo muy inicial. El análisis de IRENA indica también que las mejoras de eficiencia energética están significativamente rezagadas con respecto a los niveles requeridos. La tasa de mejora anual de la intensidad energética (energía por unidad de valor) es del 1,7%, lejos de los objetivos del 3,3% para 2030 y 2,8% para 2050, lo que indica la necesidad de intensificar los esfuerzos en este ámbito. Las inversiones asociadas a la eficiencia energética, aunque avanzan, requerirían sextuplicar los 295 mil millones de dólares anuales actuales para el año 2030.

En suma, el análisis que realiza IRENA indica que la transición energética no se encuentra en la senda compatible con el escenario de 1,5°. Este diagnóstico es compartido por otras instituciones, como la Agencia Internacional de la Energía (IEA, 2023c). En consecuencia, y máxime dadas todas las dificultades que son consustanciales a la enorme transformación tecnológica y económica que se desea acometer en el proceso de descarbonización, es razonable el empleo de técnicas de análisis que faciliten esa transición.

A ese respecto, el uso de datos masivos ha transformado la gestión energética (Jiang et al., 2016) al introducir un enfoque proactivo, predictivo y prescriptivo no solo en el mantenimiento y en las operaciones de las infraestructuras energéticas, sino en el diseño y en la innovación de soluciones y materiales con Inteligencia Artificial (IA) generativa (Fuhr y Sumpter, 2022) que eran inimaginables hasta ahora (Böcking et al., 2024). Por ejemplo, a través del empleo de algoritmos de aprendizaje automático se logra prever con precisión posibles fallos y anomalías en equipos de generación, como aerogeneradores, paneles solares, plantas hidroeléctricas o nucleares. Estas técnicas avanzadas no solo previenen interrupciones no planificadas, sino que optimizan los programas de mantenimiento, minimizando el impacto de las averías, reduciendo costos operativos y mejorando la eficiencia en la generación de energía, ya que la detección de fallos se acompaña de modelos predictivos de generación y demanda, pudiendo actuar en el momento idóneo. La detección temprana de pérdidas de rendimiento impulsa la producción, incrementando la rentabilidad de los proyectos de inversión y la energía disponible.

La gestión energética eficiente se enfoca a casos de uso de datos que aumentan la generación, especialmente en instalaciones renovables, y garantizan la seguridad de la red y el equilibrio del consumo para evitar picos de demanda (Mostafa et al., 2022). Esto incluye operaciones inteligentes que diseñan estrategias para maximizar la producción energética, como acciones ante efectos de sombreado en paneles solares, así como estrategias para gestionar la carga y descarga de baterías eléctricas y reducir la factura energética de los edificios. Asimismo, la predicción de la demanda eléctrica y térmica/refrigeración, así como de la generación renovable futura, permite mejorar las ofertas de venta y compra de energía y alimentar los sistemas de optimización antes mencionados. Al mismo tiempo, la mayor complejidad en la gestión del sistema eléctrico que aporta el autoconsumo de una parte de la energía generada por parte de empresas y hogares hace necesario disponer de datos y modelos que posibiliten una correcta integración en la red de la generación distribuida. En ese contexto, el uso de bombas de calor (Bloess et al., 2018) y el uso de las baterías de los vehículos eléctricos como elementos de almacenamiento y aporte

de energía de nuevo en la red (Alsharif et al., 2021) pueden también ayudar en esta gestión compleja de las redes.

Todas las cuestiones señaladas anteriormente serán analizadas con más detalle en apartados posteriores de este trabajo, aportando diversos casos de uso y el contexto al que responden. Interesa aquí resaltar que la gran evolución en la IA, que está transformando completamente nuestra sociedad y que se ha acelerado en los años más recientes, no hubiese sido posible sin el Big Data (BD) y las tecnologías que lo habilitan. Son diversos los factores clave que han impulsado esta evolución.

En primer lugar, el BD ha permitido la recopilación y almacenamiento masivo de datos provenientes de diversas fuentes como redes sociales o internet en general, dispositivos IoT (Internet of Things) o IIoT (Industrial IoT) (Ge et al., 2018; Tekin et al., 2024), registros financieros y blockchain (Afzal et al., 2022), entre otros. Esto ha proporcionado conjuntos más grandes y variados de datos para entrenar y validar modelos, lo que facilita la construcción de algoritmos más precisos y sofisticados.

En segundo lugar, los avances en las tecnologías de almacenamiento y procesamiento de datos, como la computación en la nube (tanto nubes públicas como privadas) y el desarrollo de hardware especializado, (desde GPUs hasta dispositivos de *edge computing* que permiten llevar la IA a los propios dispositivos; Himeur et al., 2024), han permitido manejar y procesar grandes volúmenes de datos de manera más eficiente y rápida. Esto ha posibilitado el análisis y la extracción de información valiosa de conjuntos de datos enormes.

En tercer lugar, parte del impulso al BD ha venido de la necesidad del desarrollo y la mejora de algoritmos de aprendizaje automático más complejos. Ello incluye a las redes neuronales profundas y a los algoritmos de procesamiento de lenguaje natural (NLP) (Sharma et al., 2023) como puede ser GPT o LLaMa (Touvron et al., 2023), entre otros. Estos algoritmos pueden manejar grandes cantidades de datos y extraer patrones más sutiles y complejos que los modelos tradicionales. En ese sentido, el abaratamiento del almacenamiento, la mayor capacidad de cómputo y el desarrollo de algoritmos más complejos van de la mano.

En cuarto lugar, los algoritmos de optimización (Baños et al., 2011) han experimentado avances significativos, permitiendo la mejora de la precisión y eficiencia de los modelos. Esto incluye técnicas como la optimización estocástica, algoritmos genéticos (Akachukwu et al., 2014) o la optimización bayesiana (Borunda et al., 2016), entre otras, que han permitido el ajuste más fino de los modelos a los datos. Por último, la capacidad de obtener y procesar datos en tiempo real ha impulsado el desarrollo de algoritmos y modelos que pueden adaptarse dinámicamente a cambios y actualizaciones continuas en los datos, permitiendo una toma de decisiones más ágil y precisa.

En conjunto, cabe afirmar que el BD ha generado un entorno propicio para el desarrollo de modelos y algoritmos más avanzados, permitiendo la extracción de información más detallada, la creación de predicciones más precisas y la generación de valor en diversas industrias y campos de aplicación. Antes de avanzar en los distintos ámbitos de aplicación en el sector energético, el siguiente apartado realiza una revisión de las distintas técnicas y algoritmos utilizados, algunos de ellos ya citados brevemente en los párrafos anteriores.

3. Un repaso a las técnicas y algoritmos

Las metodologías existentes, y que se han aplicado extensivamente en el contexto del sector energético, pueden clasificarse en los siguientes grupos: técnicas clásicas de series temporales (Kaur et al., 2023),

técnicas de aprendizaje supervisado y no supervisado (Alloghani et al., 2020), aprendizaje por refuerzo (Wang et al., 2020), algoritmos de optimización y *deep learning*. Dentro de este último grupo se puede incluir el uso de técnicas de *computer vision*, análisis de audio y las inteligencias artificiales generativas, como podrían ser aquellas relacionadas con grandes modelos (LLM) para el procesamiento del lenguaje natural (NLP) entre otras.

A continuación, se realiza una breve descripción de estas técnicas avanzadas de ciencia de datos y cómo el BD ha permitido llegar al estadio en el que nos encontramos hoy en día, en el que algunos investigadores reclaman que estamos más cerca de conseguir una inteligencia artificial más general, si bien, conviene señalarlo, este es un aspecto muy controvertido (Triguero et al., 2024). Esta inteligencia artificial más general tendría la peculiaridad de la intermodalidad, es decir, sería capaz de procesar datos de diferente índole (texto, números, imágenes, audio, etc.) pudiendo combinar el tratamiento de una serie temporal, bien a partir de sus datos numéricos en una tabla o bien en una gráfica. En suma, sería capaz de mostrar, de forma similar a como lo hacemos los humanos, qué fenómenos explican el comportamiento de dicha serie.

La tecnología o algoritmo que ha permitido avanzar en la IA son las redes neuronales artificiales (ANN). En puridad, no se trata de una aproximación novedosa, pues la primera publicación en este campo es del año 1958 (Rosenblatt, 1958), en correspondencia con la evolución de la electrónica y los primeros ordenadores en esos años. Sin embargo, la evolución de esta tecnología encontró diversas limitaciones relacionadas con la capacidad de cómputo y con la disponibilidad de datos de entrenamiento. Las redes neuronales, cuyo funcionamiento se explica posteriormente con más detalle, son un artefacto matemático de aprendizaje supervisado y por tanto requieren de un conjunto de datos relativamente amplio, más amplio cuanto más complejo sea el proceso que se desea modelizar.

Las ANN requieren una elevada capacidad de cómputo, ya que se entrenan mediante un proceso iterativo de ajuste de los parámetros internos. De ese modo, en sus inicios el entrenamiento de una red neuronal compleja podría necesitar cientos de años para completarse. El hecho de no poder entrenar ANN llevó a la inteligencia artificial a su primer “invierno”, con un periodo de letargo en el que su estudio tuvo poco interés y del que no se salió hasta el trabajo de Rumelhart et al. (1986) sobre la técnica de back propagation, que permitía entrenar las ANN de una forma eficiente. A este avance se unió el incesante progreso en la capacidad de cómputo en la década de los años noventa que, combinado con el despliegue de Internet en la segunda mitad de esa década, creó el caldo de cultivo perfecto para el renacer de la tecnología. A ese respecto, Internet facilitó la compartición de grandes volúmenes de datos entre usuarios y entre comunidades científicas; no solo datos numéricos sino, también, texto e imágenes. En este momento puede comenzar a hablarse propiamente de Big Data, ya que se empezaba a tener de forma conjunta y accesible grandes volúmenes de información que podía utilizarse para entrenar las ANN. Todo ello permitió la experimentación y la creación de ANN crecientemente complejas.

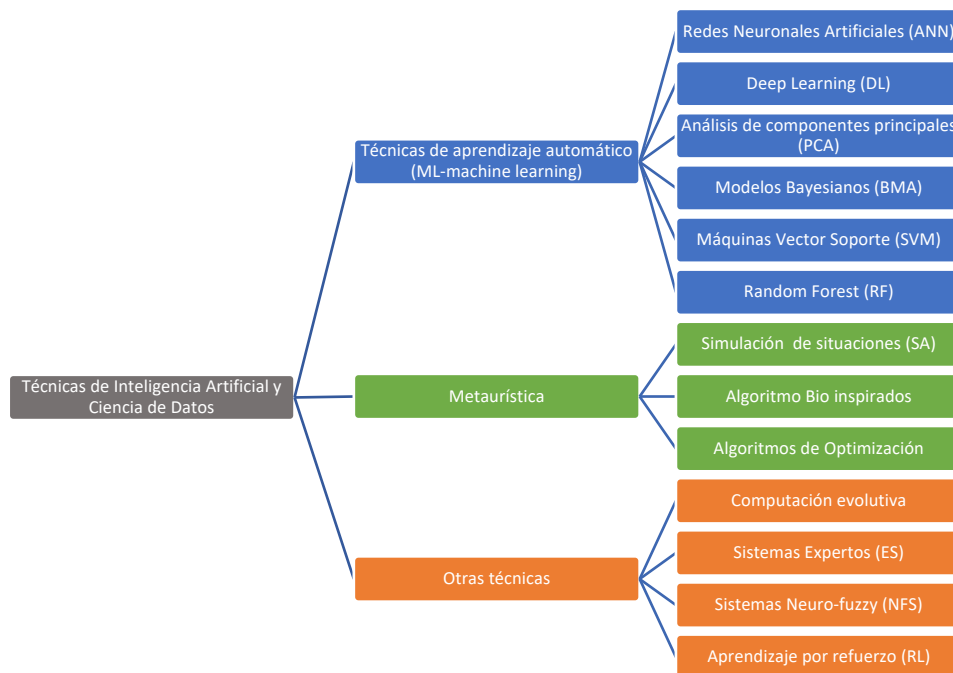
Fue a principios de la década pasada cuando se presentó un nuevo concepto de red neuronal que incluía un procesado previo de los datos para extraer sus patrones más relevantes (Krizhevsky et al., 2012). Se trata de las redes neuronales convolucionales (CNN), utilizadas principalmente para procesado de imagen (Krichen, 2023). Estas redes incluyen un procesado previo de la imagen mediante una serie de capas de convolución que extraen las principales características de cada píxel en función de otros píxeles cercanos. Es decir, se aumenta la información de un dato en función de otros datos próximos. Este uso del contexto ha sido determinante para llegar a la revolución actual de la IA.

El siguiente gran avance en la IA se produjo en 2015 con la aparición de las redes neuronales adversarias (GAN por las siglas en inglés de *Generative Adversarial Neural Networks*) (Gui et al., 2023). Una

combinación de dos redes permitía imitar la capacidad de crear e imaginar del ser humano, es decir, podíamos generar datos ficticios completamente nuevos, siendo especialmente utilizadas para la generación de imágenes.

El último gran avance, ha sido el uso de los “*transformers*” (Vaswani et al., 2023). De la misma forma en la que las capas convolucionales obtenían contexto en las imágenes, los “*transformers*” dan a cada palabra información de contexto en los textos. Esta transformación ha hecho posible el entrenamiento de redes más complejas capaces de procesar mucho mejor los documentos, lo que combinado con las técnicas de BD ha permitido entrenar gigantescos modelos de procesamiento de lenguaje natural (LLM) como son el bien conocido GPT de OpenAI o LLaMA de Facebook, entre otros. Estos modelos son capaces de procesar y argumentar de la misma forma en la que lo hacemos los humanos. Estos modelos han evolucionado y ya no son solo capaces de procesar texto, sino también imágenes, datos numéricos o audio, dando lugar a los grandes modelos multimodales (LMM) (Yin et al., 2024).

Figura 1.- Clasificación de técnicas y algoritmos basados en inteligencia artificial



Fuente: Elaborado a partir de Bennagi et al. (2024).

Los avances descritos no solo han conducido a una auténtica explosión en las publicaciones científicas aplicadas al campo de la energía, sino además han llevado a una gran variedad de soluciones tecnológicas que facilitan los progresos en el nuevo escenario energético orientado a luchar contra el calentamiento global. Sin embargo, no todo son redes neuronales artificiales en el ámbito del desarrollo de soluciones *data driven* aplicadas al sector energético. La Figura 1 ofrece una visión completa de las diferentes técnicas que se han venido aplicando. Como puede observarse, las técnicas de IA se han agrupado en algoritmos de aprendizaje automático (ML – *machine learning*), algoritmos metaheurísticos (donde se han incluido los algoritmos de optimización) y, finalmente, otras técnicas, donde se incluye el aprendizaje por refuerzo y los sistemas expertos entre otros. Del mismo modo, los algoritmos de ML pueden subdividirse en algoritmos de aprendizaje supervisado o no supervisado; incluso algunos autores

incluirían en este grupo los algoritmos de aprendizaje por refuerzo. A continuación, se describe el funcionamiento general de estas técnicas y algoritmos, adelantando algunas aplicaciones en el ámbito energético.

Aprendizaje supervisado

Las técnicas de aprendizaje supervisado pueden definirse como un enfoque en la IA y la ciencia de datos basado en el entrenamiento de modelos utilizando un conjunto de datos que incluye características (variables independientes) y sus correspondientes resultados deseados (variable dependiente o etiquetas). El modelo aprende a realizar predicciones o a clasificar nuevos datos basándose en la relación existente entre las características y los resultados conocidos del conjunto de datos de entrenamiento. En consecuencia, este tipo de modelos precisan un conjunto de datos previamente etiquetados, o bien un registro de eventos pasados para aprender de ellos. Por ejemplo, un algoritmo de aprendizaje automático puede establecer patrones de consumo energético en viviendas, pudiendo predecir que se va a tener un mayor consumo en refrigeración los días que más calor hace e, incluso, en qué medida se va a producir dicho incremento del consumo en función de la temperatura. También puede aprender de los hábitos de consumo identificando que en una determinada región se produce un pico de consumo a las ocho de la tarde, cuando se vuelve a los hogares y hay un aumento de tareas domésticas.

A partir de registros históricos de consumo, los algoritmos de aprendizaje supervisado son capaces de extraer patrones y estimar el valor de una determinada variable (consumo de una vivienda, por ejemplo) en función de otras explicativas (temperatura, hora del día, día de la semana, etc.). Esta capacidad de modelizar y predecir los consumos de cada vivienda permite, por ejemplo, gestionar de forma más segura y eficiente microrredes y facilita la integración de renovables. Entre las técnicas de aprendizaje supervisado destacan las Máquinas de Vectores de Soporte (SVM), los árboles de regresión y árboles de decisión (RF - *random forest*) y las redes neuronales artificiales (ANN). A continuación, se realiza una breve descripción de cada uno de ellos.

Las SVM son algoritmos utilizados principalmente para clasificación y regresión. La idea central de las SVM es encontrar el hiperplano óptimo que maximiza el margen entre diferentes clases en un espacio multidimensional. Por tanto, esta metodología tiene un enfoque de estudio geométrico del espacio (n dimensional) en el que se sitúan los datos. En el contexto de clasificación, cuando se trata de datos linealmente separables, las SVM buscan el hiperplano (una especie de línea en dos dimensiones, un plano en tres dimensiones y un hipervolumen en espacios de mayor dimensión) que mejor separa las clases, manteniendo la distancia máxima entre los puntos más cercanos de cada clase, conocidos como vectores de soporte. Sin embargo, cuando los datos no son linealmente separables en su espacio original, las SVM utilizan el concepto de "*kernel*" para mapear los datos a un espacio de mayor dimensión donde puedan ser separables linealmente. Esto permite encontrar un hiperplano en este nuevo espacio dimensional donde se puedan distinguir claramente las clases.

Por ejemplo, supóngase que se desea predecir la demanda eléctrica futura en una determinada región o un grupo poblacional (Dong et al., 2005). Para lograrlo, se recopila un conjunto de datos históricos que incluyen variables como la hora del día, la temperatura, la temporada del año, días festivos o eventos especiales, entre otros, y todos esos datos se relacionan con la demanda de energía eléctrica. Aplicando la metodología de SVM, se buscaría el límite geométrico que defina la mejor separación de los datos históricos de demanda eléctrica en función de las variables explicativas mencionadas. El modelo aprendería a distinguir los patrones y relaciones entre estas variables para prever con precisión la demanda eléctrica en momentos futuros.

El empleo de un "*kernel*" sería útil si los datos no fueran linealmente separables en su forma original (Zendehboudi et al., 2018). Por ejemplo, si la relación entre la temperatura y la demanda de energía no sigue una tendencia lineal, las SVM podrían mapear estos datos a un espacio de mayor dimensión donde se puedan separar de manera más clara y lineal, permitiendo así que el modelo haga predicciones más precisas. Estos pronósticos pueden utilizarse en la planificación de la generación y distribución de energía, ayudando a los proveedores a anticipar y satisfacer las necesidades de consumo de manera más eficiente.

Otro conjunto de algoritmos de aprendizaje automático supervisado son los árboles de regresión, *Random Forest* (como conjunto de varios submodelos) y las técnicas de *Bagging* y *Boosting* (Wu et al., 2018). Generalmente estas técnicas van asociadas a modelos tipo *ensembled* (Massaoudi et al., 2021), que son conjuntos de modelos más pequeños y especializados que colaboran entre sí, como se detalla más adelante. Los árboles de regresión se utilizan, como su propio nombre anuncia, para problemas de regresión, donde se busca predecir valores numéricos continuos. Estos árboles se construyen dividiendo el conjunto de datos en nodos utilizando diversas reglas de división basadas en características, con el objetivo de minimizar el error de predicción en las hojas del árbol. Por el contrario, *Random Forest* es un conjunto de árboles de decisión que trabajan en paralelo y realizan predicciones combinando las decisiones de múltiples árboles individuales. Cada árbol se entrena con una muestra aleatoria del conjunto de datos y una selección aleatoria de características. Posteriormente, las predicciones de cada árbol se promedian o se toma la moda para obtener la predicción final, reduciendo así el sobreajuste y mejorando la precisión.

En relación con estos modelos, se dispone también de las técnicas de *Bagging* y *Boosting*. *Bagging* (*Bootstrap Aggregating*) es una técnica de ensamblado que consiste en entrenar múltiples modelos independientes utilizando diferentes subconjuntos de datos generados por remuestreo (*bootstrap*) del conjunto de datos original. Posteriormente, se combinan las predicciones de estos modelos mediante promedio o votación para mejorar la precisión y reducir la varianza del modelo final. Por el contrario, *Boosting* es otra técnica de ensamblado que combina múltiples modelos débiles en un modelo más robusto. A diferencia de *Bagging*, *Boosting* entrena iterativamente modelos secuenciales, donde cada nuevo modelo se enfoca en corregir los errores del modelo anterior. Los modelos débiles se ponderan de acuerdo con su rendimiento, y se combina su salida para mejorar la precisión general del modelo. Uno de los algoritmos más populares en este ámbito es el *XGBoost* (Yazhari Kermani et al., 2024), que ha sido ampliamente utilizado por su robustez y fiabilidad. En definitiva, ambas técnicas, *Bagging* y *Boosting*, se utilizan para reducir el sobreajuste y mejorar la precisión de los modelos de aprendizaje automático al combinar múltiples modelos más simples para formar un modelo más complejo y generalizable.

La tecnología de aprendizaje supervisado que ha transformado completamente el escenario actual de la IA son las Redes Neuronales Artificiales (ANN), ya que están detrás de la inmensa mayoría de los algoritmos de *Deep Learning*. Las ANN son un modelo computacional inspirado en el funcionamiento del cerebro humano. Están compuestas por un conjunto interconectado de unidades llamadas neuronas artificiales o nodos, organizadas en varias capas, con una capa de entrada, una o más capas ocultas y una capa de salida. Cada neurona recibe una o varias entradas, las combina mediante una función de activación y genera una salida que se transmite a las neuronas de la siguiente capa. Durante el proceso de entrenamiento, las conexiones entre las neuronas, denominadas pesos, se ajustan de manera iterativa para minimizar una función de pérdida, lo que permite que la red neuronal aprenda a realizar tareas específicas, como clasificación, regresión o reconocimiento de patrones.

Hay una gran variedad de arquitecturas diferentes de redes neuronales. Las más relevantes son las redes neuronales recurrentes, las convolucionales, con memoria y las basadas en transformadores. Las redes neuronales recurrentes (RNN) están diseñadas para trabajar con datos secuenciales o temporales, capaces de capturar y procesar información de entrada con dependencias temporales gracias a la retroalimentación de sus conexiones internas. Las redes neuronales convolucionales (CNN) son un tipo de red neuronal especializada en procesar y extraer patrones de datos estructurados en forma de matrices, como imágenes, a través del uso de capas convolucionales que aplican filtros para detectar características locales y jerárquicas en la información de entrada. Las redes neuronales con memoria (LSTM, o *Long Short-Term Memory*) son un tipo de arquitectura de red neuronal recurrente diseñada para aprender dependencias a largo plazo en datos secuenciales y están estructuradas con unidades de memoria que permiten retener y actualizar información a lo largo del tiempo, lo que las hace eficientes en el procesamiento de secuencias largas y la captura de dependencias temporales a diferentes escalas. Por último, las redes basadas en el uso de transformadores, como las *Temporal Fusion Transformers* (TFT), son modelos diseñados para la predicción de series temporales (Lim et al., 2021) combinando elementos llamados “transformadores” o modelos basados en atención, que están detrás del famoso modelo del ChatGPT, con componentes específicos para el análisis y la predicción de datos temporales. Son mucho más efectivas que las RNN ya que permiten analizar la secuencia temporal a la vez y en computación paralela.

En relación con los transformadores, estos se presentaron en el año 2017 en el influyente trabajo “*Attention is all you need*” (Vaswani et al., 2017), convirtiéndose desde entonces en una arquitectura fundamental en el procesamiento de datos secuenciales, como texto, series temporales y datos estructurados. Introducidos inicialmente en el contexto del procesamiento de lenguaje natural (NLP), los transformadores revolucionaron la forma en que los modelos secuenciales comprenden y generan datos. Su innovación clave radica en la “atención”, un mecanismo que permite a los modelos enfocarse en partes específicas de la secuencia, sin importar la distancia entre ellas. Los transformadores eliminaron la necesidad de las arquitecturas recurrentes (como las citadas RNN, LSTM, etc.), que tenían dificultades con problemas como la dependencia a largo plazo y el procesamiento paralelo debido a su naturaleza secuencial.

Es importante mencionar que esta tecnología de *Deep Learning*, íntimamente relacionada con el uso de las redes neuronales (al ser un conjunto de ellas) y las técnicas de BD, ha demostrado tres grandes conjuntos de beneficios. En primer lugar, la paralelización en el cómputo al procesar entradas de modo simultáneo, lo que mejora la eficiencia computacional en comparación con las arquitecturas recurrentes. En segundo lugar, la atención permite a los transformadores capturar relaciones entre elementos distantes en la secuencia, superando la limitación de las RN en la captura de dependencias a largo plazo. Por último, son adaptables a diversas tareas, desde tareas de clasificación de texto hasta traducción automática, generación de texto y modelado de series temporales.

Todo ello ha permitido desarrollar arquitecturas que actualmente se están utilizando ampliamente en diferentes soluciones en el sector energético. Un ejemplo son los (TFT que son, como se ha señalado, una variante especializada de transformadores diseñada específicamente al análisis y predicción de series temporales. Los TFT combinan elementos de las LSTM con la potencia de atención de los transformadores y pueden capturar patrones temporales complejos, modelar relaciones entre múltiples series temporales y realizar predicciones precisas sobre el futuro de una secuencia temporal. Los TFTs también incorporan módulos de atención para capturar relaciones entre diferentes variables en series temporales multivariadas, aprovechando la información tanto de los patrones temporales como de las interacciones entre múltiples series. Esto les permite realizar pronósticos precisos incluso en entornos con

múltiples variables temporales interrelacionadas, como pronósticos de demanda, finanzas, clima y más. En suma, su capacidad para modelar y predecir patrones complejos en series temporales las convierte en herramientas poderosas en el campo del análisis de datos y la predicción en series temporales.

Aprendizaje no supervisado y algoritmos de optimización

A diferencia de los algoritmos de aprendizaje supervisado, los de aprendizaje no supervisado se utilizan para explorar y extraer patrones, estructuras o información inherente de conjuntos de datos no etiquetados. Los datos no etiquetados son observaciones en un conjunto de datos donde sólo se tienen las características o atributos de las instancias, sin información adicional sobre su clase o valor de destino. Es decir, estos datos consisten únicamente en las variables de entrada (características) y no incluyen una variable de salida (etiqueta) que se pueda utilizar como referencia para el entrenamiento de un modelo predictivo supervisado. Su objetivo principal es el de encontrar patrones ocultos o estructuras intrínsecas en los datos que puedan ser útiles para el análisis, la segmentación, la comprensión de características latentes o para reducir la dimensionalidad de los datos. Estos algoritmos trabajan con la premisa de que los datos no etiquetados tienen cierta organización interna que puede ser descubierta y utilizada para extraer información que pueda ser posteriormente explotada.

El algoritmo de *K-means* es uno de los métodos más utilizados para realizar *clustering* de datos. Un proceso de *clustering* consiste en la realización de agrupaciones de los datos en conjuntos con ciertas características comunes desconocidas a priori por el científico de datos. Su objetivo principal es agrupar un conjunto de datos en "*k*" grupos o *clusters*, donde cada punto de datos pertenece al *cluster* cuyo centroide (punto central) está más cercano a él. El algoritmo funciona de la siguiente forma. Inicialmente se seleccionan "*k*" centroides iniciales (puntos o vectores representativos) en el espacio de características. Estos centroides pueden ser puntos aleatorios del conjunto de datos o pueden ser elegidos estratégicamente. Posteriormente se calcula la distancia entre cada punto de datos y los centroides y se asigna cada punto al *cluster* representado por el centroide más cercano en función de la distancia (generalmente distancia euclidiana). Se continúa recalculando los centroides de los *clusters* como el promedio de todos los puntos asignados a ese *cluster*. Los nuevos centroides representarán el centro de masa de los puntos que pertenecen a cada *cluster*. Este proceso se repite hasta que se cumpla algún criterio de detención. El algoritmo convergerá cuando los centroides de los *clusters* no cambien significativamente entre iteraciones o cuando se alcance un número predefinido de iteraciones.

Por último, otro importante grupo de algoritmos, que se han mencionado anteriormente por su vinculación con el BD, son los algoritmos de optimización. A diferencia de los algoritmos de aprendizaje automático, que están más relacionados con grandes volúmenes de datos, los algoritmos de optimización están más ligados a la resolución de problemas con un gran número de configuraciones o soluciones posibles a un problema. Con ese fin, los algoritmos de optimización son métodos computacionales que buscan encontrar la mejor solución posible para un problema, ya sea minimizando o maximizando una función objetivo, considerando restricciones específicas. Estos algoritmos son utilizados en una amplia variedad de soluciones en el sector energético. Su objetivo es encontrar la solución óptima, aunque no necesariamente la solución perfecta, debido a la complejidad intrínseca de muchos problemas en el mundo real.

Uno de los algoritmos de optimización más conocido es el de Algoritmos Genéticos (GA). Éstos están bio-inspirados, y de ahí su nombre, en cómo se produce la evolución de las mutaciones del ADN consiguiendo que una especie se adapte mejor a su entorno. Los GA trabajan con una población de soluciones potenciales y estas soluciones son "mejoradas" a lo largo de generaciones aplicando operadores como mutación, cruce y selección, imitando el proceso de selección natural. Partiendo de una

gran variedad de configuraciones posibles para la solución de un problema, el algoritmo va combinando modificaciones de los diferentes parámetros, y en cada iteración selecciona las configuraciones que mejores resultados ofrecen, combinándolas para crear la mejor configuración que cumpla con las restricciones establecidas. En el ámbito energético, el objetivo puede ser minimizar el coste de explotación de una instalación renovable, o definir la configuración óptima de diferentes recursos renovables con el objetivo de minimizar el coste nivelado de la energía (LCOE por sus siglas en inglés²) del sistema.

4. Aplicaciones con datos masivos en generación, demanda y almacenamiento de electricidad

Hasta ahora se han proporcionado algunas ideas básicas sobre el contexto actual del BD y de las principales técnicas y algoritmos utilizados para explotar estos grandes volúmenes de datos. El objeto de este apartado es profundizar en cómo la aplicación de estas técnicas está permitiendo desarrollar una amplia variedad de soluciones en el ámbito energético, y más específicamente en el eléctrico. Al mismo tiempo, se ofrece intuición sobre cuáles son las cuestiones de mayor interés en los distintos campos de aplicación y cómo la aplicación de esas técnicas puede ayudar a avanzar en los retos planteados.

La previsión de la generación renovable

Hay una literatura empírica muy abundante sobre la predicción de la generación eléctrica renovable y sobre los precios de equilibrio en el mercado eléctrico. Ambas cuestiones están estrechamente vinculadas, porque la generación renovable es la que está sujeta a mayores incertidumbres y, en ese sentido, influye de modo más notable sobre el equilibrio del mercado, tanto en términos de las plantas de generación que entran a formar parte de ese equilibrio (ofertas infra marginales) como del precio marginal de equilibrio alcanzado en el mismo. Para ello es conveniente recordar que la generación eléctrica proviene de dos grandes grupos de centrales: de carácter gestionable y no gestionable. Las primeras incluyen las centrales térmicas (ciclos combinados, carbón y derivados del petróleo³), cuya generación permite suplir el hueco de demanda dejado por las centrales de generación renovable y cuyos costes están marcados por el coste del input primario, que en todos los casos es un hidrocarburo (gas⁴, carbón o petróleo). Esto es así porque el mercado eléctrico, tanto en Europa como en otras zonas del mundo, funciona bajo el principio de la orden de mérito. Ello no es más que la configuración de la oferta mediante la ordenación de las ofertas presentadas por las plantas de generación al mercado en función de su precio (euros por MWh), permitiendo minimizar así el coste conjunto de la generación en cada momento del tiempo. De modo natural, las plantas térmicas cuyo input primario es un hidrocarburo tendrán costes marginales más altos

² El LCOE (*Levelized Cost Of Energy*) integra los costes fijos (inversión en activos) y los costes variables (operación y mantenimiento). En el caso de la electricidad, se expresan en términos de MWh, para lo que se utilizan aproximaciones de los costes descontados y de la energía producida a la largo de la vida útil de los activos. Ello permite, además, una comparación sencilla con el precio medio de la electricidad que la instalación prevé obtener por su energía producida.

³ Debe señalarse que en España está a punto de finalizar el funcionamiento de las centrales térmicas que utilizan carbón y solo hay centrales que utilizan derivados del petróleo en los sistemas no peninsulares (Canarias, Baleares, Ceuta y Melilla). El mix eléctrico es muy heterogéneo entre los distintos Estados miembros de la Unión Europea.

⁴ Un caso especial son las centrales de cogeneración, que en su mayor parte utilizan gas natural como input primario (como las centrales de ciclo combinado) y que ofertan la totalidad de su generación al mercado, si bien aprovechan el calor generado en esas instalaciones para su uso en el proceso industrial asociado. En España, esas centrales se encuentran bajo un régimen retributivo específico, lo que debe ser tenido en cuenta a la hora de interpretar sus ofertas al mercado.

que las que hacen uso de un input primario renovable⁵. Hay algunas plantas de generación renovable que son también gestionables. El caso más significativo es el de las plantas de generación hidroeléctrica que disponen de embalse. En este caso, el titular de la planta realiza su oferta realizando una valoración del recurso primario (agua) que depende de su coste de oportunidad, determinado en el mercado diario por la tecnología que fija el precio marginal (por ejemplo, un ciclo combinado) o por las expectativas de precios futuros. También son gestionables otras plantas de generación renovable como las que utilizan biomasa o residuos que, en ese sentido, se comportan de modo similar a un ciclo combinado, teniendo en cuenta el coste variable de generación derivado del input primario consumido.

En contraste con las tecnologías anteriores, las centrales renovables no gestionables producen en función del recurso primario disponible en ese instante y no pueden aplazar la decisión de generación a un momento posterior. Ese recurso primario puede ser el sol o el viento, pero también el agua para el caso de las centrales hidroeléctricas fluyentes, esto es, que operan sin embalse. Estos generadores renovables entran con muy alta probabilidad en la casación del mercado porque actúan como precio aceptante, introduciendo en su oferta de venta precios muy bajos o cercanos a cero ya que cuentan con costes variables de generación muy reducidos. Naturalmente, la cuantía ofertada por estas plantas es muy variada porque depende de un input primario con controlable. En consecuencia, como se ha señalado, la predicción de la generación renovable no gestionable es clave para el resultado del mercado, en precios y cantidades, y por ende para la operación de todos los generadores, sean estos renovables o no.

La abundancia de parámetros que pueden condicionar la previsión de la generación renovable hace que este haya sido desde hace tiempo un campo de aplicación de las técnicas de BD. En particular, muchos trabajos han entrenado modelos de ML con múltiples datos meteorológicos (radiación, humedad, temperatura, velocidad del viento) para la predicción de la generación solar con distintos grados de granularidad geográfica. Los trabajos de Hong y Fan (2016) y Kuster et al. (2017) ofrecen una amplia revisión de las técnicas de predicción de la generación, encontrando este último que los modelos ML son más frecuentes en las aplicaciones de previsión a corto plazo, mientras que los análisis de regresión son más frecuentes en las previsiones a largo plazo. Un ejemplo más reciente y centrado en la generación renovable es el trabajo de Balal et al. (2023), en el que se utilizan distintas técnicas de ML para la predicción de generación fotovoltaica, cuyos resultados se evalúan posteriormente utilizando métricas de MSE y R^2 . Sus resultados señalan la mejoría en la capacidad predictiva asociada a RF y LSTM. Este trabajo continúa, a su vez, una larga lista de investigaciones que han utilizado distintas técnicas de ML para la predicción de generación fotovoltaica, revisados en Ahmed et al. (2020). Naturalmente, dada la relevancia que la radiación solar tiene para la predicción de la generación fotovoltaica, estos trabajos se enlazan con una literatura paralela que precisamente tiene ese objetivo y sobre los que Voyant et al. (2017) realizan una revisión.

Del mismo modo, es también muy abundante la literatura para la predicción de la generación eólica⁶. Los trabajos de Maldonado-Correa et al. (2019) y de Zhao et al. (2022) ofrecen un panorama sobre la misma. Como en el caso de la generación fotovoltaica, algunos trabajos integran modelos de predicción meteorológica con modelos supervisados de generación eólica (Kosovic et al., 2020). Además, debe señalarse que a estos estudios se están sumando otros en los que se evalúa el impacto del cambio climático

⁵ Un caso particular se refiere a la generación nuclear, que es también térmica pero con muy baja capacidad de gestionabilidad y, por tanto, con una producción prácticamente constante. Salvo circunstancias extraordinarias, los cambios en la generación se refieren a las paradas por motivos de recarga, pero estas siguen un calendario programado.

⁶ Un ámbito de aplicación adicional se refiere a la generación hidráulica. Véase, por ejemplo, Ehteram et al. (2021) para una revisión y comparación de resultados.

y del calentamiento global en la generación renovable. Rodríguez-López et al. (2020) ofrecen una aplicación al campo de la generación eólica, poniendo de manifiesto cómo el uso de redes neuronales, en combinación con reglas de lógica difusa, permiten modelizar adecuadamente la curva de potencia de las turbinas eólicas, esto es, la relación no lineal entre la electricidad generada y el viento disponible, y cómo esa relación cambia con la temperatura. Este análisis se amplía a otros casos de estudio en Rodríguez-López (2022).

Las previsiones de demanda y oferta para el equilibrio del sistema

El funcionamiento seguro del sistema energético hace imprescindible contar con previsiones muy precisas de la demanda y de la oferta disponible con distintos horizontes temporales. Esto resulta especialmente crítico en el caso del sector eléctrico, donde la operativa del sistema requiere satisfacer unos parámetros específicos de los que no puede desviarse (por ejemplo, la frecuencia) o, si produce un desvío, que este se sitúe en una banda de fluctuación muy estrecha. De hecho, es función primordial del operador del sistema eléctrico (OS) garantizar un equilibrio constante entre la oferta y la demanda inherente a ese funcionamiento. El OS realiza previsiones diarias de la demanda, así como de la generación solar y eólica, todas ellas disponibles en su página web (<https://www.esios.ree.es/es>)⁷. Esas previsiones son básicas para que pueda anticiparse a cualquier eventualidad relacionada con la gestión y seguridad del suministro. Al mismo tiempo, los participantes en el mercado también realizan previsiones sobre esa generación porque, como se ha señalado, de las mismas pueden derivarse cambios en sus estrategias de generación y de precios ofertados.

A partir de la previsión que el OS hace de esas variables y de los desajustes que se puedan poner de manifiesto a corto plazo, este requerirá a su vez la disponibilidad de diversos servicios de provisión de energía para su utilización en distintos plazos temporales. Habitualmente esos servicios de ajuste son provistos por distintos tipos de generadores que están habilitados para ello⁸, si bien también puede restablecerse el equilibrio mediante la ejecución de órdenes de disminución de la demanda a los consumidores finales que están dispuestos a ofrecer ese servicio, obteniendo a cambio la correspondiente remuneración (Duman et al., 2023). Obviamente, la provisión de esos servicios de ajuste genera un coste extra para el sistema, coste que es además creciente cuando la mayor penetración de generación renovable no se ve acompañada de más entrada de almacenamiento que actúa desplazando generación de unos momentos a otros del tiempo para aprovechar los márgenes que le proporciona el arbitraje de precios. Sin embargo, desde el punto de vista de los oferentes, esos servicios de ajuste les ofrecen ingresos adicionales, de ahí que haya interés también por parte de ellos en su predicción. Giovanelli et al. (2018) y Canizes et al. (2013) ofrecen sendas aplicaciones de redes neuronales en esta área.

Los mercados eléctricos ofrecen señales de precios a distintos plazos, si bien habitualmente una parte sustancial del volumen de transacciones se realiza en el mercado diario, que en la península ibérica es gestionado por OMIE. Hasta ahora la ventana temporal es en el mercado diario ha sido horaria, si bien

⁷ Si se desea información europea, puede consultarse la plataforma de transparencia de Entso-e (<https://transparency.entsoe.eu/>)

⁸ Los servicios de ajuste o de balance permiten al OS variar de forma rápida la energía inyectada en el sistema, al alza o la baja, haciendo frente a variaciones no esperadas de la oferta (por ejemplo, desconexión o caída de la producción de algún grupo de generación de gran tamaño) o de la demanda en un plazo de tiempo muy corto, cuya activación puede ir desde pocos segundos hasta treinta minutos. En España, los servicios de balance son los de regulación secundaria, terciaria y reemplazamiento de reserva. Estos servicios se retribuyen mediante mecanismos de mercado y, en todos los casos, requieren haber superado por parte del ofertante unas pruebas de habilitación. En el caso del gas natural, el equilibrio entre oferta y demanda no es tan exigente como en el caso del sector eléctrico, por la propia capacidad de almacenamiento de las redes de gas, de modo que los servicios de balance tienen un horizonte diario (día de gas, que no coincide con el día natural)..

pasará a ser cuarto horaria (periodos de quince minutos) a partir del año 2025. Esta es una tendencia de acercamiento de los mercados diarios al tiempo real que ya se ha producido en prácticamente la totalidad de los mercados eléctricos europeos. A ese respecto, debe recordarse que el proceso de subastas de los mercados diarios europeos es simultáneo y se efectúa mediante un algoritmo común (Euphemia). El objeto es lograr el máximo acoplamiento posible de los distintos mercados o zonas de precios a través del uso eficiente de las interconexiones, de modo que, en cada momento la electricidad fluya desde los mercados o zonas de precios con menores precios a aquellos con mayores precios. Este mercado horario se complementa con mercados de carácter intradiario (en formatos de subastas y mercado continuo), también de ámbito europeo y, como se ha señalado, con otros de mercados y servicios de ajuste a plazos más cortos.

El análisis de corto plazo sobre la oferta disponible y la demanda a satisfacer se complementa con un análisis de previsión para la cobertura del sistema en los que la ventana temporal es más amplia⁹. Así, la previsión de cobertura que los operadores europeos del sector eléctrico están obligados a realizar utiliza una ventana temporal de diez años. Ese análisis combina la previsión determinista de la oferta (capacidad disponible, basada en escenarios y paradas programadas) y la demanda (basada también en escenarios) con elementos de incertidumbre en ambos lados. Fundamentalmente, incertidumbre en relación con las condiciones meteorológicas, que afectan fundamentalmente a la demanda, y con la generación renovable disponible, cuya modelización se combina con el papel creciente que juega el almacenamiento. Los análisis de cobertura deterministas, que han sido tradicionales en la planificación energética, se basaban en considerar un escenario muy pesimista y contrastar si, con condiciones meteorológicas desfavorables, podía garantizarse un determinado umbral de exceso de generación sobre la demanda (por ejemplo, un 5%) en todo momento. Se considera que los análisis probabilistas pueden ofrecer mejores soluciones que los deterministas, especialmente en un contexto donde tanto la generación como la demanda van a ser cada vez más volátiles. La mayor volatilidad de la generación es obvia en un contexto de creciente penetración de generación renovable, pero también hay una creciente volatilidad en la demanda debido a la electrificación de nuevos usos. Por ejemplo, la mayor penetración de la bomba de calor, con consumos asociados dependientes de las temperaturas, o la mayor penetración de la movilidad eléctrica, con pautas de consumo marcadas por la intensidad de la movilidad y las características de los puntos de carga utilizados (oficinas, hogares, electrolineras).

La predicción de los precios

La predicción de los precios de la electricidad ha sido un campo de estudio recurrente en el ámbito de la economía de la energía. Muchos estudios han abordado tradicionalmente este asunto mediante la utilización de diversas técnicas econométricas que tienen en consideración múltiples características que afectan a la oferta y la demanda, tales como las condiciones meteorológicas, temperaturas, estacionalidad, etc. En ese ámbito, el trabajo de Weron (2006) es un buen ejemplo de la aplicación de distintos modelos estadísticos para la predicción de los precios de la electricidad. Sin embargo, y como ya se ha señalado, la presencia progresiva de más generación renovable en el mix eléctrico genera mayor volatilidad de los precios y, consecuentemente, mayor dificultad en su predicción tanto en el corto como en el largo plazo. De hecho, una vez transitada la crisis de precios energéticos en la Unión Europea que surgió en relación con los problemas de suministro de gas tras la invasión rusa de Ucrania, se han

⁹ Esos análisis se llevan a cabo con metodologías estandarizadas recogidas en el *European Resource Adequacy Assessment*, elaboradas por la red europea de operadores de las redes eléctricas (ENTSO-E) y actualizadas con carácter anual. Los escenarios para los distintos análisis son comunes para las redes de electricidad y gas y, a su vez, son la base del plan de desarrollo decenal de las redes (Ten-Year Network Development Plans, TYNDPs).

comenzado a poner de manifiesto los problemas de canibalización (bajos precios, incluyendo precios negativos) que ya se anticipaban por la entrada masiva de capacidad de generación renovable. A su vez, la volatilidad de los precios fuerza a los participantes a mejorar su gestión de riesgos, para lo que las mejoras en la predicción son un instrumento necesario. La propia crisis de precios en 2021-2023 puso de manifiesto cómo una inadecuada gestión de esos riesgos acentuó los problemas de competitividad para muchas empresas con consumo eléctrico intensivo o llevó a la quiebra a algunas empresas comercializadoras.

En ese contexto inherentemente más complejo, se ha extendido la aplicación de técnicas y algoritmos de IA en el ámbito de la predicción de precios de la electricidad. Aunque ya había aplicaciones hace más de dos décadas (por ejemplo, Shahidehpour et al., 2002), muchas de ellas se concentran en años más recientes. El interés se centra, en la mayoría de los casos, en los precios horarios en el mercado diario (*day ahead*), aunque la escala de los datos (horarios, diarios, semanales, ...) es un aspecto básico que se debe considerar en la interpretación de los distintos trabajos. La medida de precisión puede ser la media del porcentaje de error entre el valor real y el predicho (MAPE), un valor escalado del error (MASE) o diversas medidas de normalización alternativas (Weron, 2014). Los distintos algoritmos de redes neuronales son los más utilizados, si bien algunos con arquitecturas más sencillas (Mandal et al., 2006) que otros (Yamin et al., 2004) y, en algunos casos, combinando distintos tipos de técnicas (Luo et al., 2018; Heidarpanah et al., 2023).

El almacenamiento

Un aspecto adicional de interés y de creciente relevancia se refiere al almacenamiento que, al desplazar energía de unos momentos a otros del tiempo, afecta al equilibrio del mercado. El almacenamiento tradicional en el sector eléctrico ha sido mediante centrales que bombean y turbinan agua entre dos balsas o depósitos emplazadas a distinta altura. La gestión de esas centrales de bombeo debe tener en cuenta el momento óptimo para realizar ambas operativas, que implican consumir energía de la red (fase de bombeo) y verter energía a la red (fase de turbinación), decisiones que básicamente dependen de la previsión que el titular de la instalación hace de la curva de precios. A su vez, esa operativa influye sobre el precio del mercado, pues se trata de generación hidráulica que entra en las mismas condiciones que cualquier otra generación gestionable.

Sin embargo, las necesidades de almacenamiento son mucho mayores en un sistema con mayor penetración de generación renovable, que es a lo que se dirigen todos los sistemas eléctricos, y ello por dos conjuntos de razones. En primer lugar, porque, al traspasar energía de unos momentos a otros, el almacenamiento reduce los problemas de bajos precios en el mercado asociados a la propia entrada de generación renovable (con bajos costes medios de producción), que condicionan los incentivos a la entrada para nueva capacidad de generación renovable. En conexión con ello, también reduce los problemas de vertidos asociados a la generación renovable, consistentes en aquellas situaciones por las que una planta de generación renovable debe cesar su producción, tras una consigna del OS, por un problema de congestión de la red o un desequilibrio con la demanda. En segundo lugar, la generación renovable eólica y fotovoltaica no tienen las mismas propiedades que la generación térmica en términos de contribución a la estabilidad de la red (control de frecuencia, por ejemplo), propiedades que sí puede proporcionar el almacenamiento. Por lo tanto, el almacenamiento actúa de modo complementario a la penetración de renovables y permite su avance, con las consiguientes consecuencias positivas en términos de descarbonización (Jafari et al, 2022).

Las necesidades de aumento del almacenamiento no pueden ser suplidas con un aumento del bombeo, que se enfrenta con dificultades y plazos temporales de autorización administrativa muy largos, además

de estar sujeto a las limitaciones físicas de la disponibilidad de emplazamientos y una evolución de la curva de costes medios nivelados de la energía creciente (IRENA, 2023b). Por lo tanto, gran parte del aumento requerido de almacenamiento vendrá dado por la instalación de baterías, que cuentan con dos grandes ventajas. Por un lado, una curva de costes medios nivelados rápidamente decreciente (BNEF, 2024). Por otro lado, algunas ventajas técnicas frente a otras soluciones, como la posibilidad de ofrecer tiempos de respuesta muy rápidos. Por el contrario, las baterías sufren un proceso de desgaste rápido por la repetición de procesos de carga/descarga y tienen una capacidad de almacenamiento que permite desplazar energía entre horas del día, pero no entre periodos amplios de tiempo (por ejemplo, entre estaciones).

La creciente integración de baterías, muy evidente en sistemas eléctricos en Estados Unidos (California¹⁰, Texas), pero también en algunos países europeos (particularmente en el Reino Unido), implica gestionar las decisiones sobre los momentos óptimos de carga y descarga, y por supuesto esas decisiones afectan al equilibrio de los mercados en distintos momentos del tiempo. Se han desarrollado numerosos algoritmos que tratan de minimizar los costes de operación de las instalaciones de almacenamiento, lo que implica combinar predicciones del comportamiento del sistema instalado de baterías con predicciones del comportamiento del equilibrio de la demanda y oferta en el mercado. Esto es así porque una parte importante, aunque no la única, de los ingresos a percibir por los titulares de las instalaciones de almacenamiento provienen del arbitraje en los mercados diarios e intradiarios de electricidad. Añadidos a estos, los titulares también tienen mucho interés en modelizar el comportamiento de los mercados de servicios de ajuste, ya que estos proporcionarán otra parte de los ingresos a percibir (Varela Soares et al., 2023). En esta modelización debe tenerse muy en cuenta el distinto grado de liquidez entre los distintos mercados. En particular, el hecho de que la energía asociada a los servicios de ajuste sea pequeña en relación con los mercados de ajuste conduce a que la entrada de baterías puede canibalizar muy rápidamente los mayores precios que normalmente se observan en los mercados de servicio de ajuste en relación con los mercados diarios e intradiarios.

Debe señalarse que este análisis sobre gestión de las baterías (Battery Management System) utilizando técnicas de IA se ha desarrollado también en otras actividades que hacen uso de ellas, particularmente en el sector del automóvil. Por ejemplo, Khawaja et al. (2023) utilizan seis algoritmos de ML para analizar su rendimiento en relación con su pasado y futuro anticipado (estado de salud) así como sobre la cantidad de energía asociada a un ciclo de carga y descarga (estado de carga). El análisis se hace con baterías de iones de litio, que son también las empleadas hasta ahora en el ámbito del sector eléctrico. Sin embargo, una cuestión que debe también tenerse en cuenta es que el entorno tecnológico en el que se desarrollan estas actividades está sujeto a un acelerado proceso de cambio tecnológico, difícilmente integrable en modelos que predicen mediante su entrenamiento con datos del pasado. Así, en el caso del sector eléctrico es previsible que a mediados de la próxima década la tecnología de iones de sodio pueda desplazar a las de iones de litio, lo que implicaría diferencias apreciables en términos de costes, vida útil y características energéticas (densidad, energía específica y capacidad de carga rápida, entre otras). En Kebede et al. (2022) se ofrece un análisis comparativo sobre un amplio rango de tecnologías de almacenamiento estacionario disponibles en la actualidad.

¹⁰ El caso de California es paradigmático de la expansión en la capacidad de almacenamiento en baterías, pasando de una capacidad instalada inferior a 1 GW en 2019 a más de 10 GW instalados en 2024. De ese modo, ya en algunos momentos puntuales de este último año la electricidad inyectada en la red por parte de las baterías se convirtió en el primer suministrador por tipología de activos, por delante por ejemplo de la solar o la térmica con gas.

Por último, debe señalarse que parte de los excedentes de energía eléctrica que no puedan ser introducidos en la red, bien por falta de demanda o bien por problemas de gestión de la red, son susceptibles de ser utilizados en la generación de hidrógeno (Ge et al., 2024). En la Unión Europea, el desarrollo regulatorio sobre esta cuestión es muy reciente y, de hecho, la generación actual de hidrógeno verde es simbólica. Sin embargo, hay amplias expectativas sobre su desarrollo futuro, con una creciente literatura sobre las estrategias óptimas de dimensionamiento de la capacidad y generación (Modu et al., 2023).

La agregación de la demanda de los hogares y la generación distribuida

El equilibrio constante entre oferta y demanda, que es consustancial al funcionamiento del sector eléctrico, ha pivotado tradicionalmente por el lado de la oferta. Sin embargo, naturalmente, el equilibrio también se puede conseguir haciendo variar la demanda. En este caso, la solución tradicional ha sido la de disponer de un conjunto de empresas consumidoras de gran tamaño que, a cambio de una remuneración, están dispuestas a reducir de forma importante su consumo ante una señal del operador del sistema en un plazo de tiempo reducido. Ese era el papel que jugaba en España el ya desaparecido servicio de interrumpibilidad, y que actualmente juega el servicio de respuesta activa de la demanda. Este servicio es activado por el OS cuando necesita hacer uso de él y se otorga mediante un sistema de subasta. Una característica relevante es que, como ocurre en otros países, las empresas participantes deben ofrecer una potencia mínima que deja fuera a los consumidores de pequeño tamaño, particularmente a los hogares. La única posibilidad de que los pequeños consumidores ofrezcan un servicio de este tipo es a través de la figura de un agregador. Esta figura fue introducida en la UE por la Directiva 2019/944, de 5 de junio de 2019, sobre normas comunes para el mercado interior de la electricidad, estando pendiente aún su desarrollo reglamentario. El agregador, que podrá ser independiente del comercializador, combinará los consumos de pequeños consumidores o instalaciones de almacenamiento, ofreciendo la capacidad de modificar la energía consumida u ofertada a cambio de una remuneración (Saviuc et al., 2011).

Esta nueva figura de flexibilidad en el sistema eléctrico requiere una monitorización estrecha de los consumidores participantes por parte del agregador, así como una elevada capacidad de predicción de sus patrones de consumo y generación. Sin embargo, aunque la literatura empírica sobre respuesta de la demanda es amplia, no lo es la dirigida específicamente al campo de los agregadores de demanda¹¹, lo que en parte puede deberse al menor desarrollo de estos agentes en relación con otros servicios de respuesta de la demanda. Uno de los escasos trabajos desarrollados hasta este momento en este ámbito con técnicas de DL es el de Shaqour et al. (2022), donde los autores comparan los resultados obtenidos con cinco algoritmos alternativos y desarrollan un extenso ejercicio de sensibilidad.

Adicionalmente, el funcionamiento tradicional del sector eléctrico es el de un grupo de generadores centralizados (plantas de generación) cuya producción se encaminaba a los usuarios finales a través de las redes de transporte y distribución. Sin embargo, el sentido del flujo se vuelve crecientemente complejo con la extensión de la generación distribuida asociada al autoconsumo. El hecho de que muchos usuarios finales generen ahora parte de la electricidad que consumen, pero que al tiempo permanezcan conectados a la red y viertan a ella sus excedentes de generación, hace necesario predecir esos flujos, especialmente cuando comienzan a alcanzar un volumen significativo. Ello está llevando también a la aplicación de ML en la gestión de redes inteligentes que incorporan, entre otros rasgos, una alta digitalización y la integración de activos de almacenamiento tanto delante del contador (almacenamiento centralizado)

¹¹ Utilizando modelos de optimización, Iría y Soares (2019) analizan el funcionamiento de un agregador de pequeños agentes que actúan tanto como consumidores como oferentes. También, desde un punto de vista ingenieril, Rigoni et al. (2021) consideran el papel de los agregadores de demanda en un problema de flujos de energía óptimo.

como detrás del contador (almacenamiento distribuido). Un ejemplo significativo de ese enfoque es el trabajo de Ahmad et al. (2022), en el que se revisa con detalle las oportunidades que se abren con la utilización de algoritmos de ML en la gestión de las redes de distribución.

Asimismo, una característica de los sistemas de generación distribuida es su menor observabilidad en relación con la generación centralizada, para la que se dispone de una medición precisa por parte de los gestores de la red. Ello es debido a que (por ejemplo, en España) las pequeñas instalaciones de generación distribuida tienen un solo punto de conexión a la red, desde el que adquieren energía cuando la generación neta es negativa (esto es, la generación es menor que la generación propia) o la vierten a la red cuando es positiva (cuando la generación propia es mayor que el consumo). Por lo tanto, una adecuada gestión de la red requiere anticiparse a ese patrón de autoconsumo, teniendo en cuenta que solo se observa ese valor neto, y no el consumo y la generación a través de contadores separados. En este contexto, no se cuenta con series de datos largas, ya que este es un fenómeno relativamente nuevo. Además, la dificultad se acrecienta si se tiene en cuenta que es habitual que el gestor de la red no conozca con precisión la localización y tamaño de los equipos de generación distribuida. Por ello, algunos trabajos han propuesto el uso de ML sobre imágenes de satélite y aéreas para la identificación de esas instalaciones. El trabajo de Hoog et al. (2020) aporta un panorama de esa literatura. Por último, también debe señalarse aquí que diversos trabajos hacen uso del concepto de micro-red para abordar los retos de mayor complejidad e incertidumbre inducidos por la generación distribuida. López-García et al. (2020) realizan una amplia revisión de esos retos y cómo la aplicación de redes neuronales puede ayudar a tratar con ellos.

El consumo energético de los hogares

El consumo energético de los hogares es un componente muy relevante de la demanda final de energía: en torno a un 27% del total en la UE. Los principales usos energéticos son los asociados con la calefacción (un 65% en valor medio para la UE) y el calentamiento del agua para su uso en el hogar (14,5%). La principal fuente energética utilizada por los hogares es el gas natural (33,5%), seguida de la electricidad. El uso de combustibles fósiles hace que los hogares sean una importante fuente de emisiones de GEI, además de las emisiones indirectas que estos inducen por su consumo eléctrico. De hecho, se estima que un 36% de las emisiones de GEI en la UE provienen de los edificios¹². En consecuencia, la reducción drástica de las emisiones asociadas al parque de viviendas es una condición necesaria para la consecución de los objetivos de neutralidad de emisiones para mediados de este siglo.

Desde el punto de vista regulatorio, ello se ha traducido en una creciente atención sobre los distintos aspectos vinculados con la eficiencia del consumo energético de las edificaciones, con una reciente revisión de la Directiva de Eficiencia Energética de los Edificios que aumenta la ambición y el alcance de versiones anteriores. Junto a la revisión de esa Directiva se ha desarrollado un paquete normativo muy amplio, que va desde la calificación energética de los electrodomésticos a la certificación energética de las viviendas, entre otros aspectos. En ese contexto, es más acuciante la necesidad de mejorar las previsiones sobre el patrón de consumo energético de los hogares, y de los edificios en general. La predicción de ese consumo energético es compleja, ya que el consumo de cada unidad está afectada por múltiples factores como las características de la vivienda, el tipo de equipamiento disponible (por ejemplo, electrodomésticos) y sus calidades en términos de eficiencia energética, la localización geográfica (con sus efectos sobre las horas de sol disponibles, la temperatura exterior, la humedad o la intensidad del efecto estacional) o el propio patrón de demanda del hogar derivado de las actividades

¹² No todo el consumo energético de los edificios proviene de los hogares, pues una parte está asociado a negocios (oficinas, locales), cuyo consumo energético por unidad de superficie es más elevado. En su conjunto, se estima que los edificios son responsables del 40% del consumo energético en la UE.

laborales, periodos vacacionales o disponibilidad de una segunda vivienda, entre otros muchos aspectos. Precisamente dada la gran variedad de posibles variables de interés, los modelos de predicción del consumo de edificios más tradicionales, basados en métodos ingenieriles o estadísticos, han sido en gran medida desplazados por modelos basados en AI. Aunque prevalecen los modelos basados en ANN y en SVM, muchos trabajos combinan y evalúan la capacidad predictiva utilizando distintos algoritmos. De hecho, hay varias revisiones sobre la comparación de distintos modelos en este campo, como las de Bourdeau et al. (2019) y Zhao y Magoules (2012).

Sin embargo, debe señalarse que, como ocurre en otros ámbitos, aunque la instalación de medidores electrónicos en los hogares teóricamente ha mejorado de forma muy notable la capacidad de disponer de un amplísimo volumen de información, el investigador se enfrenta con dificultades para el acceso a esa información, pues esta se encuentra fuertemente regulada por motivos de protección de datos. En el mismo sentido, la información sobre el consumo eléctrico (o, en su caso, de gas) que se deriva de los contadores individuales es susceptible de mejora con la instalación de medidores individuales que permiten discernir con mayor precisión el comportamiento de los usuarios (consumos por electrodomésticos, televisión, calefacción, etc.) pero, de nuevo, es una información difícilmente disponible para el investigador¹³.

Por ello, es común que muchos trabajos empíricos se hayan focalizado en casos concretos para los que ha sido posible obtener información, particularmente en edificios o locales con usos comerciales o industriales. Por ejemplo, Reynolds et al. (2018) combinan una red neuronal con un algoritmo genético con el objetivo de predecir y optimizar el consumo energético de un edificio de oficinas en Cardiff. Además, dado el elevado peso de los consumos energéticos vinculados a la calefacción o aire acondicionado, es natural que diversos trabajos se hayan centrado en este ámbito. En particular, Jovanović et al. (2015) evalúan la demanda de calefacción de un campus universitario, combinando tres modelos de redes neuronales: *feed forward backpropagation neural network* (FFNN), *radial basis function network* (RBFN) y *adaptive neuro-fuzzy interference system* (ANFIS). Por su parte, Li et al. (2009) comparan el uso de distintos modelos de redes neuronales para la predicción de la demanda de aire acondicionado en un caso en China. En este trabajo, como en la mayoría de las ocasiones, el granulado temporal es horario, si bien también hay trabajos que realizan predicciones de carácter anual (Aydinalp-Koksal y Ugursal, 2008). De hecho, aunque no hay una caracterización estándar, resulta útil distinguir entre los análisis de predicción centrados en el corto plazo (por ejemplo, el consumo a lo largo de un periodo de 24 horas, o incluso el patrón entre días de una semana), en el medio plazo (por ejemplo, entre semanas o meses de un año) y en el largo plazo (periodos más largos del año). Debe señalarse que el objetivo y la utilidad de esos marcos temporales de análisis puede ser muy distinta. Por ejemplo, una mejora en la capacidad de predicción de la demanda horaria es muy útil para el funcionamiento del sistema eléctrico, mientras que el análisis para el despliegue de infraestructuras de red se beneficiará de un análisis centrado en el largo plazo.

Un ámbito de aplicación creciente se centra en la evaluación de los certificados de eficiencia energética de los edificios. Un trabajo relevante en este ámbito es el de Khayatian et al. (2016), quienes utilizan redes neuronales para evaluar qué indicadores asociados a dichos certificados son más adecuados para mejorar la capacidad predictiva sobre la demanda de calefacción de los hogares, haciendo uso de la disponibilidad pública de una base de datos online con una amplia representación de las características de los hogares en la región italiana de Lombardía. Este trabajo ofrece un ejemplo de cómo, en la elección

¹³ Por ejemplo, en España, el IDAE realiza una encuesta sobre el consumo energético de los hogares (SPAHOUSEC), pero solo es posible acceder a información agregada de ese estudio.

del algoritmo de minimización del error, la rapidez del algoritmo elegido es importante cuando, como en este caso, se cuenta con una base de datos de elevado tamaño. Otro trabajo en esta misma línea es el de Buratti et al. (2014). Algunos autores han fusionado también diversas fuentes de información a través de DL, incluyendo también el reconocimiento de imágenes de fachadas a través de Google Street View. En ese sentido, Sun et al. (2022) obtienen una importante ganancia sobre la capacidad predictiva de los modelos que solo se basan en las características morfológicas de las viviendas que se derivan de los certificados de eficiencia energética.

La demanda eléctrica de las empresas

Como se señaló con anterioridad, una parte de la demanda eléctrica de las empresas, en particular las comerciales y de servicios, está asociada a la demanda de calefacción. Sin embargo, otra parte importante de la demanda eléctrica empresarial proviene de actividades manufactureras electro intensivas, esto es, de empresas con alto consumo energético y en las que, en consecuencia, este consumo supone un porcentaje elevado en la estructura de costes de las empresas¹⁴. De este modo, un pequeño número de empresas puede representar un porcentaje relevante de la demanda eléctrica agregada¹⁵. La definición de lo que es una industria intensiva en energía puede ser muy variable, lo que explica que Zhang et al. (2018) señalen que las industrias intensivas en energía supusieron el 51% del consumo energético en China. En cualquier caso, la descarbonización del sector industrial es una pieza clave, y compleja, de la descarbonización global, por lo que hay una literatura creciente de aplicaciones *data driven* en este campo. En ese sentido, el trabajo citado de Zhang et al. (2018) propone un marco conceptual y revisa la literatura en este ámbito.

Adicionalmente, una parte creciente de la demanda eléctrica proviene de los centros de datos. La Agencia Internacional de la Energía (IEA, 2024) señala que el consumo energético mundial vinculado a centros de datos (incluidos los centros dedicados al cribado de criptomonedas) fue de 460 TWh en 2022, lo que representó un 2% de la demanda mundial y, por ejemplo, más del doble de la demanda española en ese año. Esa demanda proviene tanto de la actividad de computación y equipamientos de tecnologías de información (en torno a un 60%) como de la demanda de refrigeración asociada (en torno al 40%), previéndose un fortísimo crecimiento en los próximos años que, en buena parte, se deriva de las mayores necesidades de computación vinculadas a la IA. Diversos trabajos han explorado la aplicación de BD en esta área y tanto los trabajos de Kumar et al. (2022) como de Rajendra et al. (2022) ofrecen amplias revisiones. Por ejemplo, Vesa et al. (2020) integran un conjunto de modelos de redes neuronales para la determinación de la demanda de centros de datos, tanto en relación con su nivel como su flexibilidad. Este último aspecto es importante porque el objetivo último de ese trabajo es evaluar en qué medida los centros de datos podrían participar en los servicios de respuesta activa de la demanda, en los que tradicionalmente solo han participado hasta ahora las industrias electrointensivas¹⁶.

5. Aplicaciones data-driven en petróleo y gas natural

¹⁴ De hecho en España, como en otros países europeos, hay un reconocimiento regulatorio expreso a las empresas con consumo eléctrico intensivo, con consecuencias en términos de los peajes y cargos que estas pagan o en términos de la percepción de compensaciones por los efectos indirectos de los precios del CO₂ en su factura eléctrica.

¹⁵ En concreto, en España sólo 89 plantas industriales supusieron el 7% de la demanda eléctrica en 2023. Ello incluye fundamentalmente plantas siderúrgicas, químicas y vidrio, metalúrgicas y gases industriales.

¹⁶ Por ejemplo, en España ese servicio es demandado por el OS a través de subastas anuales. Los proveedores asignados tienen la obligación de reducir su potencia en la cuantía asignada durante un máximo de número de horas al año a cambio de una contraprestación, que es el precio establecido en la subasta.

Las actividades relacionadas con la generación, transporte y demanda de hidrocarburos (petróleo y gas natural) están siendo también un campo de aplicación relevante de las técnicas de análisis masivo de datos. Como es razonable, el tipo de cuestiones que se desea responder no siempre coincide con las analizadas para el caso de la electricidad, lo que se relaciona estrechamente con las diferencias inherentes a cada uno de los bienes y el distinto contexto de mercado y regulatorio en el que se desarrollan.

Sin embargo, debe señalarse que varias de las consideraciones sobre la previsión de la demanda realizadas en el apartado anterior son extensibles al caso del gas natural, donde el gestor técnico del sistema (GTS) también realiza previsiones de demanda con distintos horizontes de corto y medio plazo: intradiaria, diaria (día D+1) y mensual (M+1), previsiones que están disponibles en su web (<https://www.enagas.es/es/gestion-tecnica-sistema/>). En ellas es clave distinguir entre la demanda del sector eléctrico (centrales de generación de ciclo combinado) y el resto (demanda convencional), pues ambas tienen sus dinámicas propias, con influencias dispares de las temperaturas previstas, si se trata de un día de la semana y/o festivo, hora del ocaso y amanecer, etc. Por el contrario, no hay una previsión de la oferta pues, de hecho, es obligación del GTS disponer en todo momento del gas necesario, que en el caso de un país como España proviene en su integridad del exterior (importado por gasoductos o en forma de gas natural licuado). El gas suministrado a la red se acompaña con su consumo si bien, en términos del ajuste a muy corto plazo, las características del producto y del funcionamiento técnico del sistema gasista no requieren de una restauración tan inmediata del equilibrio entre oferta y demanda y, en consecuencia, la ventana temporal para alcanzar un equilibrio o balance es más amplia que en el caso del sector eléctrico. De ese modo, la literatura empírica se centra en el ámbito de la aplicación de técnicas data-driven al consumo de gas. Un ejemplo es el trabajo de Anagnostis et al. (2020), quienes evalúan tres aproximaciones de redes neuronales para la predicción del consumo de gas en Grecia, obteniendo que es aquella que integra también aspectos cualitativos la que obtiene mejor capacidad predictiva. Singh et al. (2023) y Gao y Shao (2021) ofrecen también aplicaciones para la predicción del consumo de gas para Estados Unidos y China, respectivamente, comparando distintos algoritmos, incluidos ANN y SVM.

En cualquier caso, la transición energética implica un desplazamiento desde el consumo de energías fósiles hacia vectores energéticos descarbonizados, que en la práctica son la electricidad generada con fuentes renovables y los gases renovables, esto es, hidrógeno renovable y biogás/biometano. De modo particular, el gas natural se considera como una energía de transición pues, si bien su uso genera emisiones, estas son menores que las asociadas al uso del petróleo o, en su caso, el carbón. Debe también tenerse en cuenta el distinto ámbito regulatorio existente en el caso del gas natural y en los hidrocarburos líquidos. El sector de gas natural es un sector fuertemente regulado en todo lo asociado con el suministro doméstico y, de hecho, comparte muchas de las características regulatorias del sector eléctrico, particularmente la desintegración vertical de actividades. Sin embargo, el sector de petróleo y derivados está plenamente liberalizado, sin retribución regulada a ninguno de los agentes que intervienen en la cadena de valor ni, por supuesto, precios regulados a los consumidores finales. En este caso no hay una red de distribución capilar hasta los consumidores, como ocurre en el caso de la electricidad y el gas, y la red a gestionar es básicamente una red troncal y no regulada de transporte de derivados.

En relación con los precios del petróleo, existe una abundante literatura empírica asociada a su predicción, que es inherentemente compleja y no lineal porque está sujeta a multitud de variables que afectan a la cotización internacional de este producto, lo que incluye su demanda y oferta futura y el hecho de que esa oferta se produzca en un entorno cartelizado. Esta complejidad hace especialmente útil la aplicación de técnicas de BD, particularmente en el ámbito de redes neuronales, para la predicción de precios del petróleo. Un trabajo ilustrativo en esta línea es el de Zhao et al. (2017), quienes utilizan como base un

modelo de redes neuronales (SDAE) combinado con un método de bootstrapping para generar conjuntos de datos que permitan entrenar el modelo base. Los autores incluyen 198 series económicas como variables exógenas y examinan la capacidad predictiva frente a modelos alternativos. Hay muchos otros trabajos en esta línea, como los de Shin et al. (2013) con técnicas de aprendizaje semi-supervisado (SSL), Mostafa y El-Masry (2016) con programación genética (GEP) o Godarzi et al. (2014) con un modelo dinámico de redes neuronales (NARX). Tiwari et al. (2024) realizan un pormenorizado análisis con once modelos alternativos y distintos horizontes temporales.

A ellos se unen aproximaciones más recientes que añaden el reconocimiento de textos vinculados a noticias o a búsquedas en Internet, especialmente valioso en un entorno donde los operadores de los mercados son muy sensibles a cualquier noticia que afecte a los riesgos de oferta y demanda en el futuro, riesgos que se desencadenan en muchas ocasiones por factores exógenos al mercado (por ejemplo, conflictos bélicos). Un ejemplo de esto último es el trabajo de Wu et al. (2021), quienes combinan el reconocimiento de noticias y Google Trends para predecir los precios medios semanales del crudo. La idea fundamental del trabajo es extraer de modo automático patrones ocultos en las noticias combinando *Convolutional neural network* (CNN) con técnicas de descomposición que tratan de dejar fuera la información no útil. El uso de minado de texto (text mining), que permite transformar texto no estructurado en un formato estructurado, es ya ampliamente utilizada en otros ámbitos¹⁷, pero se configura pues como una estrategia especialmente útil en un contexto de precios tan volátiles como el del crudo.

En el ámbito del gas, si bien históricamente su precio estuvo fuertemente indexado al precio del petróleo y las transacciones se efectuaban prácticamente en su totalidad en mercados no organizados, en las últimos dos décadas se han ido desarrollando mercados nacionales (*hubs*) donde se transacciona el gas con distintos horizontes temporales. El más relevante en la Unión Europea es *el Title Transfer Facility* (TTF) en los Países Bajos, cuyo amplio volumen de negocio justifica que proporcione la referencia de precios más seguida. En el caso de España, el Mercado Ibérico de Gas (MIBGAS) opera desde finales de 2015 y, también, ofrece diversos indicadores de precios con distintos periodos temporales. Debe señalarse que el peso de las transacciones a contado (*spot*) en las compras internacionales de gas ha ido aumentando, en parte debido al mayor peso relativo del gas suministrado en forma de gas natural licuado (GNL), especialmente tras la reducción de la demanda europea de gas ruso a partir de 2022. Los precios a los que se adquiere el gas en los mercados mayoristas se transmiten a los consumidores finales si bien, como en el caso de la electricidad, estos también pagan en sus facturas otros componentes vinculados a la retribución de las actividades reguladas. Su et al. (2019) ofrecen un ejemplo de predicción de los precios de contado del mercado de gas natural en Estados Unidos (Henry Hub), obteniendo mejor capacidad de predicción de ANN en relación con SVM, GBM y GPR. Adicionalmente, en su amplia revisión sobre la predicción de precios con modelos data-driven, Lu et al. (2021) revisan una amplia variedad de trabajos que han utilizado técnicas de AI para la predicción de los precios de la energía, incluyendo diversos trabajos para el gas y el petróleo.

Por último, en un ámbito distinto al de predicción de los precios, hay también un amplio uso de datos masivos en las actividades de exploración, perforación y producción de gas y petróleo, así como del refinado de sus derivados. La motivación fundamental de estos trabajos es ofrecer herramientas que faciliten la mejora en los niveles de eficiencia en esos ámbitos. En ese sentido, hay también un gran número de publicaciones científicas (Sircar et al., 2021; Tamba et al, 2018) que inciden en cómo la industria del petróleo y gas, enfrentada a una creciente competencia y volatilidad, busca innovar mediante la

¹⁷ Un caso típico se refiere a la predicción de demanda de los destinos turísticos a partir de las búsquedas en Internet. Véase, por ejemplo, Vecchio et al (2018).

digitalización, enfocándose en la eficiencia de producción, reducción de costes y mejora de la seguridad. La adopción de tecnologías basadas en la IA y el aprendizaje automático está transformando la cadena de valor, permitiendo el análisis de datos subsuperficiales más precisos, la detección avanzada de filtraciones de petróleo con robot, la perforación más eficiente con algoritmos de aprendizaje por refuerzo y el mantenimiento predictivo para evitar costosos tiempos de inactividad. Ejemplos notables incluyen colaboraciones como Total S.A. con Google Cloud para optimizar datos sísmicos, ExxonMobil con MIT en robótica submarina, Shell en perforación de precisión, y Aker BP con Spark Cognition para el mantenimiento predictivo.

Como es normal, la implementación de IA en estos ámbitos, como en otros, se enfrenta a desafíos en la integración de talento especializado, la calidad de los datos y colaboración abierta, pero promete aumentar significativamente la eficiencia operativa y la competitividad en el sector, tanto en el upstream (exploración, perforación, gestión de reservas) como en el downstream (refino y transporte) (Qing, 2021; Mohammadpoor y Torabi, 2020). En el upstream, se aborda la combinación de física digital de rocas (DRP) con big data e IA para obtener información física más precisa, lo que ayuda a mejorar el rendimiento de los yacimientos (Kang et al., 2022). De este modo, es posible integrar las características de las bases de datos de rocas con datos de pozos y registros, utilizando técnicas avanzadas de clasificación para identificar el potencial restante de petróleo y gas. Otras aplicaciones incluyen la mejora en la predicción de sismos inducidos por esas actividades, procesando amplios conjuntos de imágenes y de datos obtenidos en los procesos de extracción (presiones, temperaturas, composiciones, etc.) o la reducción de fugas de metano asociadas al transporte de gas natural (Wan et al., 2012).

6. Conclusiones

Del mismo modo que ocurre en otras áreas, el análisis aplicado en el ámbito de la economía de la energía se está beneficiando enormemente de la introducción de las técnicas y algoritmos relacionadas con el uso masivo de datos. Esto resulta especialmente importante al considerar los enormes retos que atraviesa este sector, incurso en un acelerado proceso de cambio tecnológico en el contexto de la transición energética. La creciente disponibilidad de datos por la intensa digitalización de las actividades energéticas, junto a la necesidad de una gestión inteligente de los mismos, hacen que el sector energético sea un entorno idóneo de aplicación. De hecho, la literatura ya disponible es amplísima y, en este trabajo, se ha tratado fundamentalmente de ofrecer intuición sobre cuáles son los principales asuntos que suscitan interés.

Una parte sustancial de esos trabajos tiene que ver con los efectos asociados a la penetración de electricidad de origen renovable y a los efectos que de esta se derivan en términos de gestión de la red, el almacenamiento, los precios o la gestión de la demanda, entre otros. La aplicación de técnicas de ML en cualquiera de estas áreas es ya muy extensa, pero también otros vectores, como el aún incipiente hidrógeno renovable, recibirán atención creciente en el futuro. Un aspecto fundamental en todos estos desarrollos se refiere a las mejoras en la capacidad predictiva que, en un entorno crecientemente complejo y volátil, proporcionan la aplicación de algoritmos de ML en relación con las aproximaciones estadísticas más tradicionales para distintos contextos, como la predicción de la demanda de energía por parte de hogares y empresas.

En suma, la transición energética, aunque imprescindible, es un proceso de largo recorrido que se enfrenta con barreras para el desarrollo e integración de nuevas soluciones tecnológicas y con elevados costes. El avance en el uso inteligente de los datos es una herramienta de enorme utilidad para la reducción de esos costes de transición por parte de todos los agentes involucrados.

Referencias

- Abuljadail, M., Khalil, A., Talwar, S., Kaur, P., 2023. Big Data Analytics and E-Governance: Actors, Opportunities, Tensions, and Applications. *Technological Forecasting and Social Change*. 193, 122612. <https://doi.org/10.1016/j.techfore.2023.122612>.
- Afzal, M., Li, J., Amin, W., Huang, Q., Umer, K., Ahmad, S. A., Ahmad, F., Raza, A., 2022. Role of Blockchain Technology in Transactive Energy Market: A Review. *Sustainable Energy Technologies and Assessments*. 53, 102646. <https://doi.org/10.1016/j.seta.2022.102646>.
- Ahmad, T., Madonski, R., Zhang, D., Huang, C., Mujeeb, A., 2022. Data-Driven Probabilistic Machine Learning in Sustainable Smart Energy/Smart Energy Systems: Key Developments, Challenges, and Future Research Opportunities in the Context of Smart Grid Paradigm. *Renewable and Sustainable Energy Reviews*. 160, 112128. <https://doi.org/10.1016/j.rser.2022.112128>.
- Ahmed, R., Sreeram, V., Mishra, Y., Arif, M. D., 2020. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renewable and Sustainable Energy Reviews*. 124, 109792. <https://doi.org/10.1016/j.rser.2020.109792>.
- Akachukwu, C.M., Aibinu, A.M., Nwohu, M.N., Salau, H.B., 2014. A Decade Survey of Engineering Applications of Genetic Algorithm in Power System Optimization. 5th International Conference on Intelligent Systems, Modelling and Simulation, Langkawi, Malaysia. 38-42. <https://doi.org/10.1109/ISMS.2014.15>.
- Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., Aljaaf, A.J., 2020. A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science, en: Berry, M., Mohamed, A., Yap, B. (eds) *Supervised and Unsupervised Learning for Data Science*. Unsupervised and Semi-Supervised Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-22475-2_1.
- Alsharif, A., Tan, C. W., Ayop, R., Dobi, A., Lau, K. Y., 2021. A Comprehensive Review of Energy Management Strategy in Vehicle-to-Grid Technology Integrated with Renewable Energy Sources. *Sustainable Energy Technologies and Assessments*. 47, 101439. <https://doi.org/10.1016/j.seta.2021.101439>.
- Anagnostis, A., Papageorgiou, E., Bochtis, D., 2020. Application of Artificial Neural Networks for Natural Gas Consumption Forecasting. *Sustainability*. 12, 6409. <https://doi.org/10.3390/su12166409>.
- Aydinalp-Koksal, M., Ugursal, V.I., 2008. Comparison of Neural Network, Conditional Demand Analysis, and Engineering Approaches for Modeling End-use Energy Consumption in the Residential Sector. *Appl.Energy*. 85(4), 271-296. <https://doi.org/10.1016/j.apenergy.2006.09.012>.
- Ba, Z., Ma, Y., Cai, J., Li, G., 2023. A Citation-Based Research Framework for Exploring Policy Diffusion: Evidence from China's New Energy Policies. *Technological Forecasting and Social Change*. 188, 122273. <https://doi.org/10.1016/j.techfore.2022.122273>.
- Balal, A., Pakzad, Y.J., Demir, A., Igene, M., Giesselmann, M., Bayne, S., 2023. Forecasting Solar Power Generation Utilizing Machine Learning Models in Lubbock. *Emerging Science Journal*. 7(4). <https://doi.org/10.28991/ESJ-2023-07-04-02>.
- Baños, R., Manzano-Agugliaro, F., Montoya, F.G., Gil, C., Alcayde, A., Gómez, J., 2011. Optimization Methods Applied to Renewable and Sustainable Energy: A Review. *Renewable and Sustainable Energy Reviews*. 15(4), 1753-1766. <https://doi.org/10.1016/j.rser.2010.12.008>.
- Bennagi, A., AlHousrya, O., Cofas, D. T., Cofas, P. A., 2024. Comprehensive Study of the Artificial Intelligence Applied in Renewable Energy. *Energy Strategy Reviews*. 54, 101446. <https://doi.org/10.1016/j.esr.2024.101446>.
- Bianchini, S., Damioli, G., Ghisetti, C., 2023. The environmental effects of the “twin” green and digital transition in European regions. *Environ Resource Econ* 84, 877–918. <https://doi.org/10.1007/s10640-022-00741-7>.
- Bloess, A., Schill, W., Zerrahn, A., 2018. Power-to-Heat for Renewable Energy Integration: A Review of Technologies, Modeling Approaches, and Flexibility Potentials. *Appl.Energy*. 212, 1611-1626. <https://doi.org/10.1016/j.apenergy.2017.12.073>.
- BNEF, 2024. 2H 2023 LCOE Update: An Uneven Recovery. <https://about.bnef.com/blog/2h-2023-lcoe-update-an-uneven-recovery/>.

- Böcking, L., Michaelis, A., Schäfermeier, B., Baier, A., Kühl, N., Körner, M.F., Nolting, L., 2024. Generative Artificial Intelligence in the Energy Sector. Fraunhofer FIT, Fraunhofer IEE and TenneT TSO GmbH. Bayreuth. https://doi.org/10.15495/EPub_UBT_00007674.
- Borunda, M., Jaramillo, O. A., Reyes, A., Ibarguengoytia, P. H., 2016. Bayesian Networks in Renewable Energy Systems: A Bibliographical Survey. *Renewable and Sustainable Energy Reviews*. 62, 32-45. <https://doi.org/10.1016/j.rser.2016.04.030>.
- Bourdeau, M., Zhai, X. q., Nefzaoui, E., Guo, X., Chatellier, P., 2019. Modeling and Forecasting Building Energy Consumption: A Review of Data-Driven Techniques. *Sustainable Cities and Society*. 48, 101533. <https://doi.org/10.1016/j.scs.2019.101533>.
- Buratti, C., Barbanera, M., Palladino, D., 2014. An Original Tool for Checking Energy Performance and Certification of Buildings by Means of Artificial Neural Networks. *Appl.Energy*. 120, 125-132. <https://doi.org/10.1016/j.apenergy.2014.01.053>.
- Canizes, B., Soares, J., Faria, P., Vale, Z., 2013. Mixed Integer Non-Linear Programming and Artificial Neural Network Based Approach to Ancillary Services Dispatch in Competitive Electricity Markets. *Appl.Energy*. 108, 261-270. <https://doi.org/10.1016/j.apenergy.2013.03.031>.
- de Hoog, J., Maetschke, S., Ilfrich, P., Kolluri, R.R., 2020. Using satellite and aerial imagery for identification of solar PV: state of the art and research opportunities. *e-Energy '20: Proceedings of the Eleventh ACM International Conference on Future Energy Systems*. 308-313. <https://doi.org/10.1145/3396851.3397681>.
- Dong, B., Cao, C., Lee, S. E., 2005. Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region. *Energy Build*. 37(5), 545-553. <https://doi.org/10.1016/j.enbuild.2004.09.009>.
- Duman, A. C., Gönül, Ö., Erden, H. S., Güler, Ö., 2023. Survey- and Simulation-Based Analysis of Residential Demand Response: Appliance use Behavior, Electricity Tariffs, Home Energy Management Systems. *Sustainable Cities and Society*. 96, 104628. <https://doi.org/10.1016/j.scs.2023.104628>.
- Ehteram, M., Banadkooki, F. B., Fai, C. M., Moslemzadeh, M., Sapitang, M., Ahmed, A. N., Irwan, D., El-Shafie, A., 2021. Optimal Operation of Multi-Reservoir Systems for Increasing Power Generation using a Seagull Optimization Algorithm and Heading Policy. *Energy Reports*. 7, 3703-3725. <https://doi.org/10.1016/j.egy.2021.06.008>.
- Entezari, A., Aslani, A., Zahedi, R., Noorollahi, Y., 2023. Artificial Intelligence and Machine Learning in Energy Systems: A Bibliographic Perspective. *Energy Strategy Reviews*. 45, 101017. <https://doi.org/10.1016/j.esr.2022.101017>.
- Fuhr, A.S., Sumpter, B.G, 2022. Deep Generative Models for Materials Discovery and Machine Learning-Accelerated Innovation. *Sec. Computational Materials Science*. 9 – 2022. <https://doi.org/10.3389/fmats.2022.865270>.
- Gao, F., Shao, X., 2021. Forecasting annual natural gas consumption via the application of a novel hybrid model. *Environ Sci Pollut Res*. 28, 21411–21424. <https://doi.org/10.1007/s11356-020-12275-w>.
- Ge, M., Bangui, H., Buhnova, B., 2018. Big Data for Internet of Things: A Survey. *Future Generation Comput.Syst*. 87, 601-614. <https://doi.org/10.1016/j.future.2018.04.053>.
- Ge, L., Zhang, B., Huang, W., Li, Y., Hou, L., Xiao, J., Mao, Z., Li, X., 2024. A Review of Hydrogen Generation, Storage, and Applications in Power System. *Journal of Energy Storage*. 75, 109307. <https://doi.org/10.1016/j.est.2023.109307>.
- Giovanelli, C., Sierla, S., Ichise, R., Vyatkin, V., 2018. Exploiting Artificial Neural Networks for the Prediction of Ancillary Energy Market Prices. *Energies*. 11(7), 1906. <https://doi.org/10.3390/en11071906>.
- Godarzi, A. A., Amiri, R. M., Talaei, A., Jamsab, T., 2014. Predicting Oil Price Movements: A Dynamic Artificial Neural Network Approach. *Energy Policy*. 68, 371-382. <https://doi.org/10.1016/j.enpol.2013.12.049>.
- Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J. 2023. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering*. 35(4), 3313-3332. <https://doi.org/10.1109/TKDE.2021.3130191>.
- Heidarpanah, M., Hooshyaripor, F., Fazeli, M., 2023. Daily Electricity Price Forecasting using Artificial Intelligence Models in the Iranian Electricity Market. *Energy*. 263, 126011. <https://doi.org/10.1016/j.energy.2022.126011>.

- Himeur, Y., Sayed, A. N., Alsalemi, A., Bensaali, F., Amira, A., 2024. Edge AI for Internet of Energy: Challenges and Perspectives. *Internet of Things*. 25, 101035. <https://doi.org/10.1016/j.iot.2023.101035>.
- Hong, T., Fan, S., 2016. Probabilistic Electric Load Forecasting: A Tutorial Review. *Int.J.Forecast*. 32(3), 914-938. <https://doi.org/10.1016/j.ijforecast.2015.11.011>.
- Huang, J., Chen, X., Zhao, X., 2024. How Digital Technology Reduces Carbon Emissions: From the Perspective of Green Innovation, Industry Upgrading, and Energy Transition. *J Knowl Econ*. <https://doi.org/10.1007/s13132-024-01895-z>.
- Huntington, H. G., 2021. Model Evaluation for Policy Insights: Reflections on the Forum Process. *Energy Policy*. 156, 112365. <https://doi.org/10.1016/j.enpol.2021.112365>.
- IEA, 2023a. Greenhouse Gas Emissions from Energy Data Explorer, Paris. <https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer>.
- IEA, 2023b. World Energy Outlook 2023, Paris. <https://www.iea.org/reports/world-energy-outlook-2023>.
- IEA, 2023c. Credible Pathways to 1.5 °C Four pillars for action in the 2020s, Paris. <https://www.iea.org/reports/credible-pathways-to-150c>.
- IEA, 2024. Electricity 2024. Analysis and forecast to 2026. <https://www.iea.org/reports/electricity-2024>.
- IRENA, 2023a. World Energy Transitions Outlook 2023. <https://www.irena.org/Digital-Report/World-Energy-Transitions-Outlook-2023>.
- IRENA, 2023b. Renewable power generation costs in 2022. <https://www.irena.org/Publications/2023/Aug/Renewable-Power-Generation-Costs-in-2022>.
- Iria, J., Soares, F., 2019. Real-Time Provision of Multiple Electricity Market Products by an Aggregator of Prosumers. *Appl.Energy*. 255, 113792. <https://doi.org/10.1016/j.apenergy.2019.113792>.
- Jafari, M., Botterud, A., Sakti, A., 2022. Decarbonizing Power Systems: A Critical Review of the Role of Energy Storage. *Renewable and Sustainable Energy Reviews*. 158, 112077. <https://doi.org/10.1016/j.rser.2022.112077>.
- Jiang, H., Wang, K., Wang, Y., Gao, M., Zhang, Y., 2016. Energy big data: A survey. *IEEE Access*. 4, 3844-3861. <https://doi.org/10.1109/ACCESS.2016.2580581>.
- Jovanović, R.Ž., Sretenović, A.A., Živković, B.D., 2015. Ensemble of various Neural Networks for Prediction of Heating Energy Consumption. *Energy Build*. 94, 189-199. <https://doi.org/10.1016/j.enbuild.2015.02.052>.
- Kang, J., Li, N., Zhao, L., Xiong, G., Wang, D., Xiong, Y., Luo, Z., 2022. Construction of Complex Digital Rock Physics Based on Full Convolution Network. *Petroleum Science*. 19(2), 651-662. <https://doi.org/10.1016/j.petsci.2021.11.018>.
- Kaur, J., Parmar, K.S., Singh, S., 2023. Autoregressive models in environmental forecasting time series: a theoretical and application review. *Environ Sci Pollut Res*. 30, 19617–19641. <https://doi.org/10.1007/s11356-023-25148-9>.
- Kebede, A.A., Kalogiannis, T., Van Mierlo, J., Bercibar, M., 2022. A Comprehensive Review of Stationary Energy Storage Devices for Large Scale Renewable Energy Sources Grid Integration. *Renewable and Sustainable Energy Reviews*. 159, 112213. <https://doi.org/10.1016/j.rser.2022.112213>.
- Khayatian, F., Sarto, L., Dall'O', G., 2016. Application of Neural Networks for Evaluating Energy Performance Certificates of Residential Buildings. *Energy Build*. 125, 45-54. <https://doi.org/10.1016/j.enbuild.2016.04.067>.
- Khawaja, Y., Shankar, N., Qiqieh, I., Alzubi, J., Alzubi, O., Nallakaruppan, M.K., Padmanaban, S., 2023. Battery Management Solutions for Li-Ion Batteries Based on Artificial Intelligence. *Ain Shams Engineering Journal*. 14(12), 102213. <https://doi.org/10.1016/j.asej.2023.102213>.
- Kosovic, B., Haupt, S.E., Adriaansen, D., Alessandrini, S., Wiener, G. et al, 2020. A comprehensive wind power forecasting system integrating artificial intelligence and numerical weather prediction. *Energies*. 13, 1372. <https://doi.org/10.3390/en13061372>.
- Krichen, M., 2023. Convolutional Neural Networks: A Survey. *Computers*. 12, 151. <https://doi.org/10.3390/computers12080151>.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012)*. <https://doi.org/10.1145/3065386>.

- Kumar, R., Kumar, S.K., Diván, M.J., 2022. Optimization of power consumption in data centers using machine learning based approaches: a review. *International Journal of Electrical and Computer Engineering*. 12, 3. <http://doi.org/10.11591/ijece.v12i3.pp3192-3203>.
- Kuster, C., Rezgui, Y., Mourshed, M., 2017. Electrical Load Forecasting Models: A Critical Systematic Review. *Sustainable Cities and Society*. 35, 257-270. <https://doi.org/10.1016/j.scs.2017.08.009>.
- Li, Q., Meng, Q., Cai, J., Yoshino, H., Mochida, A., 2009. Predicting Hourly Cooling Load in the Building: A Comparison of Support Vector Machine and Different Artificial Neural Networks. *Energy Conversion and Management*. 50(1), 90-96. <https://doi.org/10.1016/j.enconman.2008.08.033>.
- Lim, B., Arik, S. Ö, Loeff, N., Pfister, T., 2021. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int.J.Forecast.* 37(4), 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>.
- Lima, M.A., Carvalho, P., Fernández-Ramírez, L.M., Braga, A.P.S., 2020. Improving Solar Forecasting using Deep Learning and Portfolio Theory Integration. *Energy*. 195, 117016. <https://doi.org/10.1016/j.energy.2020.117016>.
- Lonergan, K. E., Suter, N., Sansavini, G., 2023. Energy Systems Modelling for just Transitions. *Energy Policy*. 183, 113791. <https://doi.org/10.1016/j.enpol.2023.113791>.
- Lopez-Garcia, T.B., Coronado-Mendoza, A., Domínguez-Navarro, J. A., 2020. Artificial Neural Networks in Microgrids: A Review. *Eng Appl Artif Intell*. 95, 103894. <https://doi.org/10.1016/j.engappai.2020.103894>.
- Lu, H., Ma, X., Ma, M., Zhu, S., 2021. Energy Price Prediction using Data-Driven Models: A Decade Review. *Computer Science Review*. 39, 100356. <https://doi.org/10.1016/j.cosrev.2020.100356>.
- Luo, X., Zhu, X., Lim, E.G., 2018. A hybrid model for short term real-time electricity price forecasting in smart grid. *Big Data Anal.* 3, 8. <https://doi.org/10.1186/s41044-018-0036-x>.
- Maldonado-Correa, J., Solano, J.C., Rojas-Moncayo, M., 2021. Wind Power Forecasting: A Systematic Literature Review. *Wind Eng.* 45(2), 413-426. <https://doi.org/10.1177/0309524X19891672>.
- Mandal, P., Senjyu, T., Funabashi, T., 2006. Neural Networks Approach to Forecast several Hour Ahead Electricity Prices and Loads in Deregulated Market. *Energy Conversion and Management*. 47(15), 2128-2142. <https://doi.org/10.1016/j.enconman.2005.12.008>.
- Massaoudi, M., Refaat, S.S., Chihi, I., Trabelsi, M., Oueslati, F.S., Abu-Rub, H., 2021. A Novel Stacked Generalization Ensemble-Based Hybrid LGBM-XGB-MLP Model for Short-Term Load Forecasting. *Energy*. 214, 118874. <https://doi.org/10.1016/j.energy.2020.118874>.
- Modu, B., Abdullah, M.P., Bukar, A. L., Hamza, M.F., 2023. A Systematic Review of Hybrid Renewable Energy Systems with Hydrogen Storage: Sizing, Optimization, and Energy Management Strategy. *Int J Hydrogen Energy*. 48(97), 38354-38373. <https://doi.org/10.1016/j.ijhydene.2023.06.126>.
- Mohammadpoor, M., Torabi, F., 2020. Big Data Analytics in Oil and Gas Industry: An Emerging Trend. *Petroleum*. 6(4), 321-328. <https://doi.org/10.1016/j.petlm.2018.11.001>.
- Mostafa, M. M. and El-Masry, A. A., 2016. Oil Price Forecasting using Gene Expression Programming and Artificial Neural Networks. *Econ.Model.* 54, 40-53. <https://doi.org/10.1016/j.econmod.2015.12.014>.
- Mostafa, N., Ramadan, H. S. M., Elfarouk, O., 2022. Renewable Energy Management in Smart Grids by using Big Data Analytics and Machine Learning. *Machine Learning with Applications*. 9, 100363. <https://doi.org/10.1016/j.mlwa.2022.100363>.
- Qing, W., 2021. Global Practice of AI and Big Data in Oil and Gas Industry, en Patrick Bangert (ed.): *Machine Learning and Data Science in the Oil and Gas Industry*. Capítulo 9, 181-210: Gulf Professional Publishing.
- Rajendra, K., Sunil, K., Khatri, M.J., 2022. Optimization of power consumption in data centers using machine learning based approaches: a review. *International Journal of Electrical and Computer Engineering*. 12, 3192-3203. <https://doi.org/10.11591/ijece.v12i3.pp3192-3203>.
- Reynolds, J., Rezgui, Y., Kwan, A., Piriou, S., 2018. A Zone-Level, Building Energy Optimisation Combining an Artificial Neural Network, a Genetic Algorithm, and Model Predictive Control. *Energy*. 151, 729-739. <https://doi.org/10.1016/j.energy.2018.03.113>.
- Rigoni, V., Flynn, D., Keane, A., 2021. Coordinating Demand Response Aggregation with LV Network Operational Constraints. *IEEE Transactions on Power Systems*. 36(2), 979-990. <https://doi.org/10.1109/TPWRS.2020.3014144>.
- Rodríguez-López, M.Á., 2015. Metodología para sistemas inteligentes de detección de mal funcionamiento en equipos. Universidad de La Rioja. <https://dialnet.unirioja.es/servlet/tesis?codigo=46488>.

- Rodríguez-López, M.Á., 2022. Climate Change Impacts on Renewable Energy Generation and Energy Generation Scenarios. <https://hdl.handle.net/20.500.14352/3953>.
- Rodríguez-López, M.Á., Cerdá, E., Rio, P.d., 2020. Modeling Wind-Turbine Power Curves: Effects of Environmental temperature on Wind Energy Generation. *Energies*. 13(18), 4941. <https://doi.org/10.3390/en13184941>.
- Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*. 65(6), 386-408. <https://doi.org/10.1037/h0042519>.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature*. 323, 533-536. <https://doi.org/10.1038/323533a0>.
- Saviuc, I., Zabala López, C., Puskás-Tompos, A., Rollert, K., Bertoldi, P., 2022. Explicit Demand Response for small end-users and independent aggregators – Status, context, enablers and barriers. Publications Office of the European Union, Luxembourg. <https://doi.org/10.2760/625919>, JRC129745.
- Shahidehpour, M., Yamin, H., Li, Z., 2002. Market operations in electric power systems: forecasting, scheduling, and risk management. John Wiley & Sons, New York.
- Shaqour, A., Ono, T., Hagishima, A., Farzaneh, H., 2022. Electrical Demand Aggregation Effects on the Performance of Deep Learning-Based Short-Term Load Forecasting of a Residential Building. *Energy and AI*. 8, 100141. <https://doi.org/10.1016/j.egyai.2022.100141>.
- Sharma, S., Malik, A., Sharma, C. et al., 2023. Adoption of industry 4.0 in different sectors: a structural review using natural language processing. *Int J Interact Des Manuf*. <https://doi.org/10.1007/s12008-023-01550-y>.
- Singh, S., Bansal, P., Hosen, M., Bansal, S. K., 2023. Forecasting Annual Natural Gas Consumption in USA: Application of Machine Learning Techniques- ANN and SVM. *Resour.Policy*. 80, 103159. <https://doi.org/10.1016/j.resourpol.2022.103159>.
- Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., Oza, H., 2021. Application of machine learning and artificial intelligence in oil and gas industry. *Petroleum Research*. 6(4), 379-391. <https://doi.org/10.1016/j.ptlrs.2021.05.009>.
- Shin, H., Hou, T., Park, K., Park, C., Choi, S., 2013. Prediction of Movement Direction in Crude Oil Prices Based on Semi-Supervised Learning. *Decis. Support Syst*. 55(1), 348-358. <https://doi.org/10.1016/j.dss.2012.11.009>.
- Su, M., Zhang, Z., Zhu, Y., Zha, D., Wen, W., 2019. Data Driven Natural Gas Spot Price Prediction Models Using Machine Learning Methods. *Energies*. 12, 1680. <https://doi.org/10.3390/en12091680>.
- Sun, M., Han, C., Nie, Q., Xu, J., Zhang, F., Zhao, Q., 2022. Understanding Building Energy Efficiency with Administrative and Emerging Urban Big Data by Deep Learning in Glasgow. *Energy Build*. 273, 112331. <https://doi.org/10.1016/j.enbuild.2022.112331>.
- Tamba, J.G., Essiane, S.N., Sapnken, E.F., Koffi, F.D., Nsouandélé, J.L., Soldo, B., Njomo, D., 2018. Forecasting natural gas: A literature survey. *Int. J. Energy Econ. Policy* 2018, 8, 216–249.
- Taniguchi-Matsuoka, A., Shimoda, Y., Sugiyama, M., Kurokawa, Y., Matoba, H., Yamasaki, T., Morikuni, T., Yamaguchi, Y., 2020. Evaluating Japan's National Greenhouse Gas Reduction Policy using a Bottom-Up Residential End-use Energy Simulation Model. *Appl.Energy*. 279, 115792. <https://doi.org/10.1016/j.apenergy.2020.115792>.
- Tekin, N., Aris, A., Acar, A., Uluagac, S., Gungor, V. C., 2024. A Review of on-Device Machine Learning for IoT: An Energy Perspective. *Ad Hoc Networks*. 153, 103348. <https://doi.org/10.1016/j.adhoc.2023.103348>.
- Tiwari, A. K., Sharma, G. D., Rao, A., Hossain, M. R., Dev, D., 2024. Unraveling the Crystal Ball: Machine Learning Models for Crude Oil and Natural Gas Volatility Forecasting. *Energy Econ*. 134, 107608. <https://doi.org/10.1016/j.eneco.2024.107608>.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A. et al., 2023. LLaMA: Open and Efficient Foundation Language Models. <https://arxiv.org/abs/2302.13971>.
- Triguero, I., Molina, D., Poyatos, J., Del Ser, J., Herrera, F., 2024. General Purpose Artificial Intelligence Systems (GPAIS): Properties, Definition, Taxonomy, Societal Implications and Responsible Governance. *Information Fusion*. 103, 102135. <https://doi.org/10.1016/j.inffus.2023.102135>.
- Varela Soares, I., Mauger, R., Santos, T., 2023. Considerations for Benefit Stacking Policies in the EU Electricity Storage Market. *Energy Policy*. 172, 113333. <https://doi.org/10.1016/j.enpol.2022.113333>.

- Vaswani, A., Shazeer, M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2023. Attention Is All You Need. <https://arxiv.org/abs/1706.03762>.
- Vecchio, P.D., Mele, G., Ndou, V., Secundo, G., 2018. Creating Value from Social Big Data: Implications for Smart Tourism Destinations. *Information Processing & Management*. 54(5), 847-860. <https://doi.org/10.1016/j.ipm.2017.10.006>.
- Vesa, A.V., Cioara, T., Anghel, I., Antal, M., Pop, C., Iancu, B., Salomie, I., Dadarlat, V.T., 2020. Energy Flexibility Prediction for Data Center Engagement in Demand Response Programs. *Sustainability*. 12, 1417. <https://doi.org/10.3390/su12041417>.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine Learning Methods for Solar Radiation Forecasting: A Review. *Renewable Energy*. 105, 569-582. <https://doi.org/10.1016/j.renene.2016.12.095>.
- Wan, J., Yu, Y., Wu, Y., Feng, R., Yu, N., 2012. Hierarchical Leak Detection and Localization Method in Natural Gas Pipeline Monitoring Sensor Networks. *Sensors*. 12(1), 189-214. <https://doi.org/10.3390/s120100189>.
- Wang, H., Liu, N., Zhang, Y. et al., 2020. Deep reinforcement learning: a survey. *Front Inform Technol Electron Eng*. 21, 1726–1744. <https://doi.org/10.1631/FITEE.1900533>.
- Weron, R., 2006. Modeling and forecasting electricity loads and prices: A statistical approach. John Wiley, England.
- Weron, R., 2014. Electricity Price Forecasting: A Review of the State-of-the-Art with a Look into the Future. *Int.J.Forecast*. 30(4), 1030-1081. <https://doi.org/10.1016/j.ijforecast.2014.08.008>.
- Wu, B., Wang, L., Lv, S., Zeng, Y., 2021. Effective Crude Oil Price Forecasting using New Text-Based and Big-Data-Driven Model. *Measurement*. 168, 108468. <https://doi.org/10.1016/j.measurement.2020.108468>.
- Wu, Z., Li, N., Peng, J., Cui, H., Liu, P., Li, H., Li, X., 2018. Using an Ensemble Machine Learning Methodology-Bagging to Predict Occupants' Thermal Comfort in Buildings. *Energy Build*. 173, 117-127. <https://doi.org/10.1016/j.enbuild.2018.05.031>.
- Yamin, H. Y., Shahidehpour, S. M., Li, Z., 2004. Adaptive Short-Term Electricity Price Forecasting using Artificial Neural Networks in the Restructured Power Markets. *International Journal of Electrical Power & Energy Systems*. 26(8), 571-581. <https://doi.org/10.1016/j.ijepes.2004.04.005>.
- Yazhari Kermani, A., Abdollahi, A., Rashidinejad, M., 2024. Cyber-Secure Energy and Flexibility Scheduling of Interconnected Local Energy Networks with Introducing an XGBoost-Assisted False Data Detection and Correction Method. *International Journal of Electrical Power & Energy Systems*. 155, 109683. <https://doi.org/10.1016/j.ijepes.2023.109683>.
- Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., 2024. Enhong Chen, A Survey on Multimodal Large Language Models. <https://arxiv.org/abs/2306.13549>.
- Zendehboudi, A., Baseer, M. A., Saidur, R., 2018. Application of Support Vector Machine Models for Forecasting Solar and Wind Energy Resources: A Review. *J.Clean.Prod*. 199, 272-285. <https://doi.org/10.1016/j.jclepro.2018.07.164>.
- Zhang, Y., Ma, S., Yang, H., Lv, J., Liu, Y., 2018. A Big Data Driven Analytical Framework for Energy-Intensive Manufacturing Industries. *J.Clean.Prod*. 197, 57-72. <https://doi.org/10.1016/j.jclepro.2018.06.170>.
- Zhao, H., Magoulès, F., 2012. A Review on the Prediction of Building Energy Consumption. *Renewable and Sustainable Energy Reviews*. 16(6), 3586-3592. <https://doi.org/10.1016/j.rser.2012.02.049>.
- Zhao, Y., Li, J., Yu, L., 2017. A Deep Learning Ensemble Approach for Crude Oil Price Forecasting. *Energy Econ*. 66, 9-16. <https://doi.org/10.1016/j.eneco.2017.05.023>.
- Zhao, E., Sun, S., Wang, S., 2022. New Developments in Wind Energy Forecasting with Artificial Intelligence and Big Data: A Scientometric Insight. *Data Science and Management*. 5(2), 84-95. <https://doi.org/10.1016/j.dsm.2022.05.002>.
- Zhou, K., Fu, C., Yang, S., 2016. Big Data Driven Smart Energy Management: From Big Data to Big Insights. *Renewable and Sustainable Energy Reviews*. 56, 215-225. <https://doi.org/10.1016/j.rser.2015.11.050>.