

Optimización del consumo energético en centros de datos mediante modelos de regresión múltiple basados en carga de procesamiento y métricas de red

Optimization of energy consumption in data centers using multiple regression models based on processing load and network metrics

Jaime David Camacho Castillo¹, Joel Geovanny Bonifaz Mendoza², Juan Sebastián Montesdeoca Guzmán³, Samanta Kassiel Narvaez Coba⁴, Alexis Xavier Nicolalde Carrasco⁵, John Omar Yugcha Guevara⁶

Cita:

Jaime David Camacho Castillo et al. (2026). Optimización del consumo energético en centros de datos mediante modelos de regresión múltiple basados en carga de procesamiento y métricas de red. *TESLA Revista Científica*, 6(2), e723. <https://doi.org/10.55204/trc.v6i2.e723>

Recibido: 06/05/2026

Revisado: 18/05/2026–08/06/2026

Corregido: 15/06/2026

Aceptado: 25/06/2026

Publicado: 08/09/2026

Licencia:

Los contenidos de este artículo están bajo una licencia Creative Commons Attribution 4.0 International (CC BY 4.0). Los autores conservan sus derechos morales y patrimoniales.

¹ Escuela Superior Politécnica de Chimborazo (ESPOCH), Riobamba, Ecuador.

²⁻⁶ Investigador externo, Ecuador.

¹ jaimed.camacho@espoch.edu.ec, ² joelbonifazmendoza@gmail.com, ³ jmontesdeoca682@gmail.com, ⁴ samykassiel21@gmail.com, ⁵ aleniccar@gmail.com, ⁶ johnyugcha0@gmail.com

¹0000-0002-9110-6585, ²0009-0008-8326-4314, ³0009-0003-3339-5264, ⁴0009-0005-5481-3837, ⁵0009-0008-5612-0534, ⁶0009-0004-2962-8181

Resumen: La digitalización global ha consolidado a los centros de datos como infraestructura crítica, aunque su consumo energético crece de manera desproporcionada y desafía la sostenibilidad. La ausencia de sensores físicos de medición en servidores virtualizados ha impulsado el desarrollo de modelos predictivos basados en software. En este estudio, se realizó una revisión sistemática siguiendo el protocolo PRISMA (2021-2026) sobre modelos de Regresión Lineal Múltiple (MLR) para estimar el consumo energético en centros de datos. Se analizaron métricas SNMP de CPU, red y memoria, aplicando técnicas de preprocesamiento como imputación, normalización y mitigación de multicolinealidad. Los modelos MLR híbridos con Gradient Boosting alcanzaron un MAE de 1.46 y un R² superior al 95%. Estos resultados demuestran que los modelos software pueden sustituir a los sensores físicos, facilitando la consolidación dinámica de máquinas virtuales y el control predictivo de HVAC. Las futuras líneas de investigación incluyen el Descubrimiento Causal (Causal ML) para identificar las causas del estrés térmico, la Regresión Simbólica adaptable a cambios de hardware y la extrapolación a redes híbridas renovables en entornos rurales de edge computing. La MLR híbrida se posiciona como un sustituto preciso de la instrumentación física, reduce el consumo inerte hasta un 10% sobre el límite ideal y permite un control climático proactivo. Es necesario integrar Causal ML y Regresión Simbólica para gestionar la estocasticidad en arquitecturas futuras.

Palabras clave: Regresión Lineal Múltiple, centros de datos, eficiencia energética, consolidación dinámica, descubrimiento causal, regresión simbólica.

Abstract: Global digitalization has established data centers as critical infrastructure, yet their energy consumption is increasing disproportionately, posing significant sustainability challenges. The absence of physical measurement sensors in virtualized servers has driven the adoption of software-based predictive models. This study conducted a systematic review using the PRISMA protocol (2021-2026) focused on Multiple Linear Regression (MLR) models for estimating energy consumption in data centers. SNMP metrics from CPU, network, and memory were analyzed, with preprocessing steps including imputation, normalization, and mitigation of multicollinearity. Hybrid MLR models incorporating Gradient Boosting achieved a mean absolute error (MAE) of 1.46 and an R² exceeding 95%. These findings demonstrate that software models can effectively replace physical sensors, supporting dynamic virtual machine consolidation and predictive HVAC control. Future research directions include the application of Causal ML to identify the etiological origins of thermal stress, Symbolic Regression adaptable to hardware changes, and extrapolation to hybrid renewable networks in rural edge computing environments. Hybrid MLR serves as a precise substitute for physical instrumentation, reduces idle consumption to within 10% of the ideal limit, and enables proactive climate control. The integration of Causal ML and Symbolic Regression is necessary to address stochasticity in future architectures.

Keywords: Multiple Linear Regression, data centers, energy efficiency, dynamic consolidation, causal discovery, symbolic regression.

INTRODUCCIÓN

La digitalización mundial, impulsada por la extensa difusión de servicios de computación en la nube, el ambiente del Internet de las Cosas (IoT) y la llegada masiva de algoritmos de aprendizaje automático (Biehl, 2023; Dudek, 2023), ha puesto a los centros de datos como uno de los elementos clave de la infraestructura tecnológica actual. Sin embargo, la creación y el manejo de estos Centros de Datos de Hiperescala (HDC) implica un gasto energético masivo que desafía las normas de sostenibilidad en diversas ubicaciones geográficas (Zhang & Liu, 2022). El requerimiento de energía mundial para estas instalaciones ha crecido con rapidez y se espera que siga subiendo sin pausa hasta el año 2030, haciendo de los centros de datos uno de los mayores consumidores de electricidad en el ámbito industrial global.

Este uso intenso de energía plantea problemas serios en relación con su impacto negativo en el medio ambiente, la producción de desechos electrónicos y las emisiones de carbono constantes, lo que hace urgente la adopción de medidas de flexibilidad conscientes del carbono y estrategias sistémicas de eficiencia (De Brito et al., 2023; Misaghian et al., 2023; Teresa Riccio Barbosa & Santiago Do Nascimento Lima, 2025). Además, en la era de los grandes datos, no gestionar este impacto de forma proactiva puede traducirse en un riesgo de estrés financiero corporativo, evidenciando la necesidad de optimización a nivel estructural (Gabrielli et al., 2026).

Para evaluar y controlar esta situación, el sector ha empezado a usar métricas de eficiencia, consolidando la Efectividad del Uso de Energía (PUE) como el parámetro más importante, cuyo comportamiento predictivo es analizado exhaustivamente mediante diversas técnicas de machine learning (Kumar et al., 2021; Muraku et al., 2024). El PUE calcula la relación total de energía eléctrica que recibe un edificio con respecto a la energía usada solo por el equipamiento informático principal. Llevar el valor del PUE a su punto óptimo de 1.0 requiere una mejora rigurosa de la infraestructura física, lo que abarca la implementación de acuerdos de compra de energía (Arellanos-Tafur et al., 2025), la evaluación con diferentes tipos de economizadores climáticos (Kim et al., 2024) y la optimización de redes híbridas renovables, como el monitoreo IoT en plantas solares en azoteas o el control predictivo de paneles (Reyes Almeida, 2021; Tenenaula Cunduri, 2025; Vergara et al., 2025).

Asimismo, demanda una reestructuración de la carga computacional subyacente y de los esquemas de refrigeración, integrando marcos de control jerárquico para sistemas HVAC y algoritmos proactivos impulsados por modelos de memoria a corto y largo plazo (LSTM) para un análisis térmico preciso (Liu et al., 2026; Wang, Zhou, et al., 2025). Esta convergencia metodológica resalta el inexorable movimiento hacia la "Computación en la Nube Verde", garantizando al mismo tiempo la reducción de violaciones en los acuerdos de nivel de servicio (Biswas et al., 2021).

Sin embargo, una de las principales dificultades para lograr esta eficiencia detallada, incluso en infraestructuras avanzadas como centros de telecomunicaciones basados en virtualización de funciones de red (NFV) o racks de computación de alto rendimiento, es la ausencia general de sensores de medición de energía integrados en cada servidor dentro de vastos ambientes virtualizados (Ennio et al., 2025; Wang et al., 2023). La instalación de hardware de medición eléctrica en tiempo real, similar al empleado en máquinas industriales bajo el paradigma de la industria 4.0 (Bautista Sánchez, 2022), resulta sumamente costosa y representa un desafío logístico crítico en escenarios de gestión multizonal inciertos (Rafique et al., 2026). A esta limitación física se añade que gran parte del software en la nube opera en hipervisores, los cuales aíslan a las Máquinas Virtuales (VM) del soporte físico, lo que limita severamente la visibilidad directa del consumo eléctrico asociado a un proceso concreto e incrementa la complejidad del descubrimiento causal en las operaciones diarias (Zapata et al., 2024).

Como respuesta a estas restricciones, la investigación científica ha convergido en el uso de herramientas basadas en software que utilizan modelos analíticos y predictivos, incluyendo el aprendizaje profundo por refuerzo, para pronosticar las cargas de trabajo y evaluar el estado energético global (Karantoumanis & Ploskas, 2020; Khan et al., 2022; Ma & Ding, 2022). Relacionando el consumo energético total de un nodo con métricas de rendimiento y metadatos de tareas, tales como los aplicados en la predicción del consumo durante el entrenamiento de GPU mediante regresión simbólica, se pueden crear medidores virtuales altamente precisos (Liao et al., 2026). Esto facilita un manejo eficiente de los recursos en la nube, emulando la rigurosidad requerida en estimaciones probabilísticas de demanda observadas en otros dominios, como las estaciones de carga de vehículos eléctricos o las estrategias de predicción térmica en edificaciones generales (Lascano et al., 2023; Ortega-Díaz et al., 2023; Tarafdar et al., 2023).

Por lo tanto, la Regresión Lineal Múltiple (MLR) se ha convertido en el eje analítico principal de las estrategias de predicción de consumo más avanzadas, demostrando ser una herramienta matemáticamente robusta para modelar desde operaciones complejas de blockchain hasta la disipación de calor en centros de datos submarinos (Cesario et al., 2025; Georgiev & Nanjappan, 2024; Jin et al., 2023; Yan & Ming, 2023). Al considerar múltiples variables a la vez, la MLR permite crear una representación multidimensional de la actividad del sistema, superando los diseños de línea base tradicionales (Crespo et al., 2021) y permitiendo integraciones innovadoras, como la combinación con técnicas de gradient boosting focalizadas en la carga del hipervisor (Afrianto et al., 2025).

La evolución de este modelado aborda además problemas estadísticos profundos mediante el uso de regresión bayesiana para mitigar la colinealidad, modelos de punto de unión (JP-MLR) para electricidad en edificios, y aproximaciones lineales ponderadas (Tricube) apoyadas en herramientas que aseguran estabilidad y distribución espacial correcta (Al-Essa et al., 2024; George et al., 2023; Guañuna et al., 2023; Rantini et al., 2026; Zheng et al., 2022). Todo ello garantiza que la infraestructura subyacente opere no solo con agilidad, sino bajo parámetros estrictos de sostenibilidad continua.

En virtud de lo expuesto, la presente revisión sistemática se conduce bajo la siguiente premisa, estructurada mediante el formato PICO:

¿Cómo se compara la precisión de los modelos MLR híbridos con la instrumentación física en la gestión de eficiencia y SLA en centros de datos?

Para responder a esta interrogante, los componentes PICO se definieron de la siguiente manera:

- **Población (P):** Centros de datos de hiperescala y entornos virtualizados.
- **Intervención (I):** Modelos de Regresión Lineal Múltiple (MLR) e híbridos.
- **Comparación (C):** Sensores de medición físicos y modelos univariados (SLR/ARIMA).
- **Resultado (O):** Precisión predictiva (MAE, RMSE, R^2), reducción de consumo inerte y optimización de métricas de eficiencia energética.

METODOLOGÍA

Para garantizar la integridad estructural y la reproducibilidad del análisis, el presente estudio se estructura adoptando los lineamientos estandarizados por el protocolo PRISMA. La estrategia metodológica se divide orgánicamente en cuatro fases principales.

Protocolo de Revisión Sistemática PRISMA

El diagrama de flujo del protocolo PRISMA aplicado constó de las clásicas cuatro etapas: Identificación, Cribado Screening, Elegibilidad e Inclusión.

• Fase 1: Identificación

Durante el primer trimestre de 2026, realizamos una búsqueda bibliográfica exhaustiva para recopilar la evidencia necesaria. Consultamos bases de datos internacionales como Scopus, IEEE Xplore, ScienceDirect, MDPI y Springer Nature, además de repositorios regionales como SciELO y Dialnet.

Para que el proceso sea reproducible y riguroso, diseñamos una ecuación de búsqueda booleana que adaptamos a la sintaxis de cada motor:

("Multiple Linear Regression" OR "MLR" OR "Regression Models") AND ("Data Center" OR "Cloud Computing" OR "Hyperscale") AND ("Energy efficiency" OR "Power consumption" OR "Predictive modeling")

Fijamos criterios de selección claros: artículos de investigación y actas de conferencias publicados entre 2021 y 2026, escritos en inglés o español y que hayan pasado por revisión de pares.

• Fase 2: Cribado (Screening)

Una vez recolectados los registros, usamos un gestor bibliográfico para limpiar la base de datos, eliminando duplicados de forma manual y automática. Luego, revisamos los títulos y resúmenes de manera independiente para descartar estudios puramente teóricos o revisiones que no presentaran una validación cuantitativa sólida vinculada a nuestra pregunta de investigación.

• Fase 3: Elegibilidad

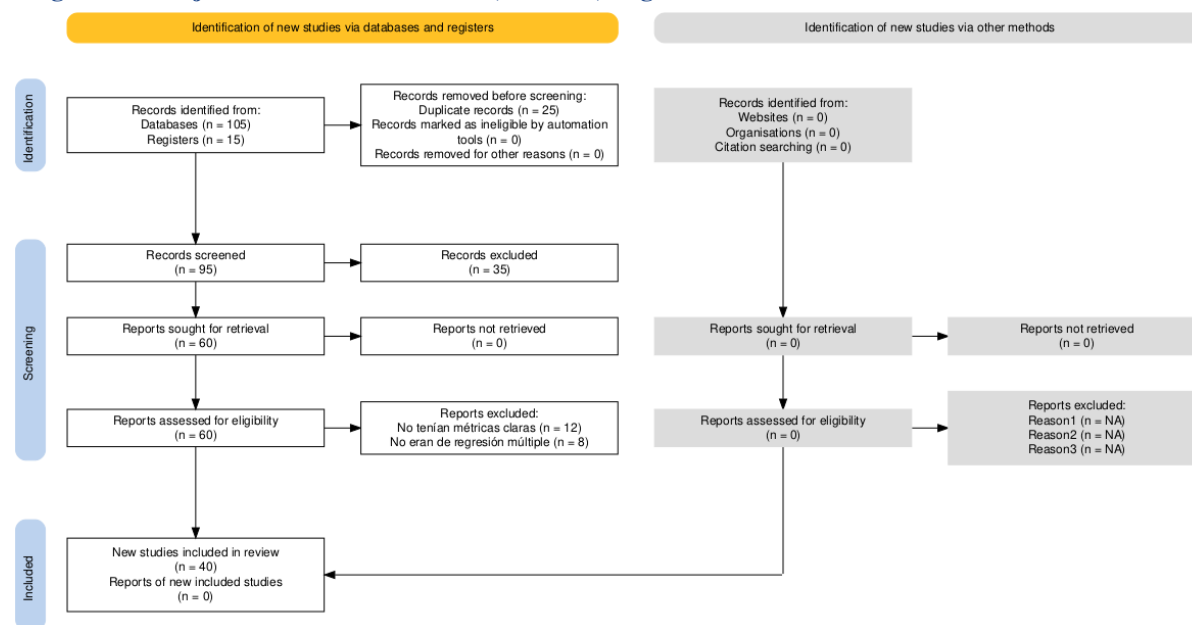
En esta etapa, analizamos el texto completo de los artículos preseleccionados. Nos enfocamos en la calidad metodológica, priorizando revistas de alto impacto con DOI y literatura regional que tuviera una aplicación clara en infraestructuras energéticas. Dejamos fuera aquellos estudios con metodologías poco claras, falta de métricas de error o un alto riesgo de sesgo.

• Fase 4: Inclusión

Finalmente, el corpus de estudio quedó conformado por 40 fuentes primarias. Con este grupo consolidado, procedimos a la extracción de datos, enfocándonos en variables críticas como los algoritmos de regresión utilizados, las arquitecturas evaluadas y las métricas de rendimiento (específicamente MAE, RMSE y R^2).

Figura 1

Diagrama de flujo PRISMA: identificación, cribado, elegibilidad e inclusión



Nota: Elaboración propia con base en el protocolo PRISMA.

Preprocesamiento de Datos de Hipervisor y Métricas de Red

El rendimiento del modelo de Regresión Lineal Múltiple depende en gran medida de la calidad de los datos de entrada. Por ello, la captura de la telemetría se realiza de forma no intrusiva mediante el Protocolo Simple de Administración de Red (SNMP), extrayendo variables clave asociadas al uso de CPU, red y memoria.

Para preparar esta información, es necesario aplicar un preprocesamiento exhaustivo. Este proceso abarca la imputación de valores faltantes, el tratamiento de valores atípicos (outliers) para evitar sesgos en los coeficientes, y la normalización de los datos utilizando métodos como Min-Max o Z-score. Finalmente, se realiza un análisis de correlación para mitigar la multicolinealidad, asegurando la selección de características ortogonales que maximicen la varianza explicada del modelo sin generar redundancia.

Modelado: Regresión Lineal Simple y Múltiple

Regresión Lineal Simple (SLR): Asume una relación unidimensional entre una variable independiente (usualmente el uso de CPU) y el consumo. Se formula como $E = \beta_0 + \beta_1 * U_c + \varepsilon$. Aunque es computacionalmente ligera, resulta insuficiente frente a cargas modernas intensivas en memoria o red.

Regresión Lineal Múltiple (MLR): Se formula como un hiperplano multidimensional: $E = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$. Las variables X_i representan el conjunto de características preprocesadas. Los coeficientes se estiman mediante Mínimos Cuadrados Ordinarios (OLS). La literatura muestra variaciones avanzadas como la Regresión Lineal Ponderada Tricube (TWLR) para priorizar observaciones recientes, o modelos híbridos en dos fases que utilizan Gradient Boosting para suavizar componentes no lineales antes de aplicar la MLR, mitigando así la volatilidad estocástica.

Mecanismos de Optimización

El propósito de la Regresión Lineal Múltiple (MLR) en este contexto es generar predicciones en tiempo real para facilitar una orquestación inteligente. Estas estimaciones sirven de base para la "Consolidación Dinámica de Máquinas Virtuales", una estrategia que concentra las cargas de trabajo en la menor cantidad posible de servidores físicos, lo que permite apagar o suspender los nodos que quedan inactivos. De forma complementaria, el Control Predictivo Basado en Modelos (MPC) regula los sistemas de enfriamiento (HVAC) anticipándose al calor que generarán los servidores. Esto ayuda a prevenir cuellos de botella térmicos y asegura que no se incumplan los Acuerdos de Nivel de Servicio (SLA).

Evaluación de Calidad y Riesgo de Sesgo.

Para garantizar la validez empírica de la revisión sistemática, evaluamos rigurosamente la calidad metodológica de los 40 artículos que conformaron el corpus final. Damos prioridad a las investigaciones que probaron sus modelos con datasets reconocidos (como PlanetLab, CloudSim o datos reales de clústeres de Google) o en entornos experimentales que pudieran ser replicados.

Por otro lado, descartamos desde las primeras fases aquellos estudios con un alto riesgo de sesgo; es decir, los que no detallaban sus métricas de error predictivo (MAE, RMSE, R^2) o que no incluían un modelo de referencia (como SLR o modelos ARIMA básicos) para demostrar que su propuesta multivariable era realmente superior. Finalmente, comprobamos que los autores reportaran de forma transparente la instrumentación utilizada a través de protocolos estándar como SNMP.

RESULTADOS Y DISCUSIÓN

Como resultado de la fase de extracción de datos, se consolidó la evidencia cuantitativa y cualitativa de las fuentes primarias. En la Tabla 1 se sintetizan los aportes empíricos y métricas de rendimiento de los estudios más representativos del corpus analizado, los cuales sustentan la viabilidad de la regresión múltiple en la estimación energética.

Tabla 1

Síntesis de la evidencia extraída de los estudios principales

Autor y Año	Modelo / Enfoque	Dataset o Métricas Utilizadas	Resultados Principales / Métricas de Error
Afrianto et al. (2025)	MLR + Gradient Boosting	Métricas SNMP (User, Inbound, System)	MAE = 1.46, RMSE = 1.50. Reducción de varianza del 86.42% frente a ARIMA.
Zheng et al. (2022)	JP-MLR (Joinpoint-MLR)	Uso eléctrico en infraestructuras	Coefficiente de determinación $R^2 = 95.77\%$.
Cesario et al. (2025)	Modelos predictivos consolidación	Telemetría del dataset PlanetLab	Consumo reducido a 1181.54 KWh (solo 10.37% sobre el límite ideal).
Biswas et al. (2021)	MLR para gestión de recursos	Entornos Green Cloud Computing	Reducción conjunta del consumo energético y de la tasa de violación de SLA.
Liao et al. (2026)	Regresión Simbólica	Metadatos de tareas en GPU	Descubrimiento de morfologías adaptables a cambios de hardware y throttling.
Yan y Ming (2023)	MLR para disipación de calor	Flujo térmico en centros submarinos	Estabilización inercial y optimización directa del indicador PUE perimetral.
Tenenaula Cunduri (2025)	Simulación sistemas híbridos	Variables análogas de energía rural	Alta precisión predictiva con un $R^2 = 0.936$.
Crespo et al. (2021)	Regresión multivariable	Línea base en plantas industriales	Fuerte correlación estadística con un $R^2 > 0.77$.

Nota: Elaboración propia con base en la información analizada.

Eficacia del Preprocesamiento y Dependencia Matemática

El estudio de Afrianto et al. (2025) validó que el uso selectivo de métricas SNMP (cargas de *User*, tráfico *Inbound*, interrupciones *System*) actúa como un sustituto correlacional perfecto de los medidores de hardware. El filtrado adaptativo de picos anómalos mediante Gradient Boosting asegura que los datos residuales cumplan las asunciones probabilísticas de la MLR.

Desempeño Predictivo de los Modelos (Evaluación de Errores)

La evaluación cuantitativa del modelo híbrido YUGO demostró un Error Absoluto Medio (MAE) de apenas 1.46 y un Error Cuadrático Medio (RMSE) de 1.50 para ventanas temporales amplias, lo que representa una reducción histórica del 86.42% en la varianza de error frente a modelos ARIMA base. Zheng et al. (2022) lograron con el modelo JP-MLR un coeficiente de determinación (R^2) del 95.77%, validando la trazabilidad de la regresión frente a métodos de "caja negra".

A nivel regional, Crespo et al. (2021) aplicaron modelos multivariados en plantas industriales alcanzando correlaciones superiores al 77% ($R^2 > 0.77$), mientras que simulaciones de Tenenaula Cunduri (2025) en sistemas híbridos lograron un $R^2 = 0.936$ empleando variables análogas. Estos estudios consolidan el consenso empírico de que la MLR multivariable trasciende fronteras y es aplicable a diversas topologías.

Impacto en la Consolidación (SLA) y Dimensión Térmica (HVAC)

Cesario et al. (2025) utilizaron el dataset de PlanetLab para probar que la consolidación predictiva de VMs suprime el consumo base inerte. Mientras un enfoque reactivo tradicional consumió 1518.44 KWh, el enfoque predictivo consumió solo 1181.54 KWh,

apenas un 10.37% por encima del límite teórico inalcanzable ideal de 1070.50 KWh. Biswas et al. (2021) complementaron esto demostrando que el modelado preciso reduce concomitantemente el consumo energético y la tasa de violación de SLAs.

En la dimensión térmica, las proyecciones anticipadas alimentan el control HVAC proactivo. Modelos como los propuestos por Yan y Ming (2023) para centros submarinos evidencian que predecir el flujo térmico estabiliza las métricas inerciales y optimiza asombrosamente el PUE perimetral.

Aplicaciones prácticas

Los resultados de este análisis se traducen en aplicaciones operativas directas. La principal ventaja es que permite reemplazar la instrumentación de hardware físico (como sensores y vatímetros perimetrales) por modelos predictivos basados en software, obteniendo la telemetría del hipervisor a través de SNMP. Con esto, los ingenieros y arquitectos de nube pueden integrar los modelos matemáticos directamente en las capas de orquestación. Así, es posible gestionar dinámicamente la consolidación de las máquinas virtuales y reducir el desperdicio de energía generado por los servidores inactivos.

Al mismo tiempo, estas predicciones facilitan un control climático (HVAC) proactivo, ya que permiten ajustar la refrigeración con horas de anticipación a los picos de carga transaccional. En cuanto a la consolidación y los Acuerdos de Nivel de Servicio (SLA), anticiparse a la demanda elimina el consumo base innecesario. Por ejemplo, estudios como el de Cesario et al. (2025) han logrado reducir el gasto energético, dejándolo apenas un 10.37% por encima del límite teórico ideal. Esto supera con creces a los enfoques reactivos y disminuye a la par las tasas de incumplimiento de servicio (Biswas et al., 2021).

De manera similar, en el ámbito térmico, estas proyecciones estabilizan las métricas inerciales y mejoran de forma considerable la efectividad del uso de la energía (PUE). Esto dota a los sistemas HVAC de la proactividad necesaria para garantizar la sostenibilidad operativa de la nube (Yan y Ming, 2023).

Futuras líneas de investigación

Aunque los modelos ensamblados actuales resuelven exitosamente gran parte de la estocasticidad, las arquitecturas ultra-densas del mañana exigen una evolución algorítmica. A continuación, se presentan las principales líneas futuras identificadas:

Descubrimiento Causal (Causal ML)

Sugerimos explorar el campo del Descubrimiento Causal combinado con la analítica predictiva. A diferencia de los modelos puramente predictivos, el Causal ML va más allá de la simple observación de patrones, permitiendo inferir relaciones direccionales de causa y efecto. Esto ayuda a mapear cómo las latencias ocultas y fluctuantes de la red afectan la temperatura de distintos componentes de la placa base. Al construir un Grafo Acíclico Dirigido (DAG) a partir de la telemetría SNMP, los algoritmos no solo podrán pronosticar la volatilidad térmica, sino también identificar su origen exacto.

Regresión Simbólica y mutación morfológica

Consideramos a la Regresión Simbólica (SR) como un enfoque fundamental para el futuro, ya que permite extraer ecuaciones matemáticas que se adaptan a los cambios en la topología del hardware y al tipo de tareas ejecutadas (como el entrenamiento intensivo en GPU). A diferencia de la MLR tradicional, la SR hace evolucionar su estructura y parámetros mediante programación genética, lo que facilita la captura de cambios bruscos de régimen, como el *throttling* térmico. Recomendamos evaluar estrategias híbridas (SR con Gradient Boosting) y neuro-simbólicas (DSR, SNAM) para unir la búsqueda evolutiva discreta con la optimización continua.

Extrapolación a redes híbridas renovables en entornos rurales

Otra línea de proyección con alto impacto regional y social es la adaptación de estos algoritmos de MLR para infraestructuras *edge* alimentadas por energías renovables. En zonas rurales o contextos educativos, como en el cantón Riobamba, proponemos desarrollar modelos de dos capas: una para estimar la generación fotovoltaica según el clima y otra para predecir la demanda eléctrica de los servidores. Esta combinación serviría para alimentar un planificador energético (*energy-aware scheduling*) que traslade las tareas no críticas a las horas de mayor luz solar, reduciendo la huella de carbono y la dependencia de la red comercial (Tenenaula Cunduri y Benítez Bravo, 2025; Vergara et al., 2025).

6.4 Modelos adaptativos y online

Para entornos que procesan metadatos de GPU en *streaming*, recomendamos implementar una programación genética incremental (iGP) con buffers FIFO y olvido adaptativo (ADWIN). Esto permite que los modelos se ajusten sobre la marcha a cambios no estacionarios en los datos sin tener que reentrenar todo desde cero. En el caso de las arquitecturas con múltiples GPUs, se podría usar una variante federada que sincronice los modelos mediante el promedio de sus configuraciones (*average pooling*), lo que reduciría la latencia para alcanzar un consenso.

CONCLUSIONES

A continuación, se detalla la jerarquización de las inferencias analíticas definitivas ordenadas desde su mayor impacto macroestructural hasta las orientaciones técnicas para futuras investigaciones:

La Regresión Lineal Múltiple (MLR) híbrida funciona como un reemplazo virtual muy efectivo para los sensores físicos, logrando reducir el error absoluto medio (MAE) hasta en un 86% en comparación con los modelos ARIMA tradicionales. Para los investigadores del área, esto confirma que las métricas de CPU y red obtenidas vía SNMP son indicadores más que suficientes para estimar el consumo de potencia con alta precisión. Visto desde otras disciplinas (como la ingeniería eléctrica o las energías renovables), este hallazgo demuestra que es totalmente viable implementar mediciones virtuales sin depender de instrumentación costosa, validando el uso de la MLR en entornos con limitaciones de hardware.

La aplicación práctica más directa de este enfoque es la orquestación predictiva en la nube. Esta técnica permite pasar los servidores inactivos a estados de suspensión profunda, logrando que el consumo energético del centro de datos se acerque a un nivel óptimo (con apenas un 10.37% de desviación sobre el límite teórico ideal). Para el público general, esto se traduce en servicios en la nube más económicos y con una menor huella de carbono. Además, demuestra que anticiparse a la demanda no

solo ahorra energía, sino que también mantiene intacto el cumplimiento de los Acuerdos de Nivel de Servicio (SLA).

Por otro lado, sincronizar las predicciones de carga computacional con la infraestructura de enfriamiento permite ajustar la ventilación y las bombas de agua antes de que ocurran los picos de temperatura. Esto elimina el retraso natural que tienen los sistemas reactivos. Para los arquitectos de centros de datos y los ingenieros de climatización (HVAC), esto representa una reducción drástica del indicador PUE (Efectividad del Uso de Energía) total.

A nivel general, la implementación de estos descubrimientos nos acerca al objetivo de una "Nube Verde" más sostenible. No obstante, dado que estas conclusiones se basan en escenarios de carga controlados, todavía se consideran preliminares para entornos altamente estocásticos. Por ello, es necesario impulsar futuros estudios que:

- 1. Integren el Descubrimiento Causal (Causal ML) para identificar el origen exacto del estrés térmico.
- 2. Apliquen Regresión Simbólica para que los modelos se adapten de forma autónoma a los cambios en el hardware.
- 3. Extrapolen estos modelos predictivos a infraestructuras edge en zonas rurales, alimentadas por redes híbridas de energía renovable.

FINANCIACIÓN

La presente investigación no contó con ningún tipo de financiamiento externo, gubernamental, institucional, becas ni privado. Fue desarrollada de manera independiente, utilizando exclusivamente los recursos personales y tecnológicos propios de los autores.

CONFLICTO DE INTERESES

Los autores declaran expresamente que no existe ningún conflicto de intereses, financiero, personal o de otra índole, que pudiera haber influido en los resultados, la interpretación algorítmica o la redacción del presente trabajo investigativo.

CONTRIBUCIÓN DE AUTORÍA

Tabla de contribución

Contribución de autoría según la taxonomía CRediT

Participar activamente en:	Camacho	Bonifaz	Montesdeoca	Narvaez	Nicolalde	Yugcha
Conceptualización	X	X	X	–	–	–
Análisis formal	X	–	X	X	X	–
Adquisición de fondos	–	–	–	–	–	–
Investigación	–	X	X	X	X	X
Metodología	X	X	–	–	X	X
Administración del proyecto	X	X	–	–	–	X
Recursos	–	X	–	X	X	X
Redacción – borrador original	X	X	X	X	X	X
Redacción – revisión y edición	X	X	X	X	X	X
Discusión de los resultados	X	X	X	X	X	X
Revisión y aprobación de la versión final	X	X	X	X	X	X

Nota: Distribución declarada conforme a la taxonomía CRediT.

REFERENCIAS

Afrianto, Y., Munadi, R., Setyorini, & Widyanto, T. (2025). *Y UGO: Enhancing Power Consumption Predictions in Data Centers Through Combined Multiple Linear Regression and Gradient Boosting Techniques on Hypervisor Load*. *IEEE Access*, 13, 210025-210039. <https://doi.org/10.1109/ACCESS.2025.3604711>

Al-Essa, L. A. & others. (2024). Bayesian regression modeling and inference of energy efficiency data: The effect of collinearity and sensitivity analysis. *Frontiers in Energy Research*. <https://doi.org/10.3389/fenrg.2024.1416126>

Arellanos-Tafur, E., Gutierrez Rupay, H., Rojas-Arquiñego, F., & Damas-Niño, R. (2025). Impact of energy efficiency on reducing consumption in data centers with power purchase agreements (PPAs). *LACCEI International Multi-Conference for Engineering, Education, and Technology*. https://laccei.org/LACCEI2025-Mexico/full-papers/Contribution_2072_final_a.pdf

Bautista Sánchez, J. V. (2022). *Sistema de medición en tiempo real y gestión del consumo energético de máquinas industriales mediante habilitadoras digitales de la industria 4.0*.

Biehl, M. (2023). *The Shallow and the Deep: A biased introduction to neural networks and old school machine learning*. University of Groningen Press. <https://doi.org/10.21827/648c59c1a467e>

Biswas, N. Kr., Banerjee, S., Biswas, U., & Ghosh, U. (2021). An approach towards development of new linear regression prediction model for reduced energy consumption and SLA violation in the domain of green cloud computing. *Sustainable Energy Technologies and Assessments*, 45, 101087. <https://doi.org/10.1016/j.seta.2021.101087>

Cesario, E., Lindia, P., Lobello, F., Vinci, A., & Capalbo, S. (2025). Enhancing energy efficiency in cloud computing through regression models: A data-driven approach with experimental validation. *Future Generation Computer Systems*, 177. <https://doi.org/10.1016/j.future.2025.108245>

Crespo, G. & others. (2021). Modelos de regresión y diseño de línea base para indicadores energéticos en una empresa siderúrgica. *Ingeniería Energética*, 42(1), 1-10.

- De Brito, J. L. R., Matai, P. H. L. D. S., & Dos Santos, M. R. (2023). Data Center e Eficiência Energética: Data Center and Energy Efficiency. *Brazilian Journal of Business*, 5(2), 786-795. <https://doi.org/10.34140/bjbv5n2-002>
- Dudek, G. (Ed.). (2023). *Applied Machine Learning*. MDPI - Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/books978-3-0365-7907-8>
- Ennio, F. T. L., Cruz, A. R. D. L., & Rodelas, N. C. (2025). JANUS: Study and analysis of the cooling and power requirements of an NFV telecommunications data center using machine learning. En *Emerging Technologies In Sustainable Innovation, Management and Development*. Routledge.
- Gabrielli, G., Melioli, A., & Bertini, F. (2026). Corporate financial distress prediction: A machine learning approach in the era of big data. *Journal of Accounting and Organizational Change*, 22(7), 31-65. <https://doi.org/10.1108/JAOC-05-2025-0166>
- George, N., K. Anoop, B., & P. Vijayan, V. (2023). Tricube Weighted Linear Regression and Interquartile for Cloud Infrastructural Resource Optimization. *Computer Systems Science and Engineering*, 45(3), 2281-2297. <https://doi.org/10.32604/csse.2023.028117>
- Georgiev, G. V., & Nanjappan, V. (2024). Predicting the Energy Consumption of Blockchain Operations: A Multiple Linear Regression Approach. *Sustainability*, 16(4), 1519. <https://doi.org/10.3390/su16041519>
- Guañuna, G. F., Granda, N. V., Cepeda, J. C., Vargas, W., & Echeverría, D. E. (2023). Estimación del Margen de Estabilidad de Voltaje Utilizando Herramientas de Aprendizaje Automático. *Revista Técnica "energía"*, 20(1), 1-8. <https://doi.org/10.37116/revistaenergia.v20.n1.2023.570>
- Jin, T. & others. (2023). Energy consumption prediction and optimization of data centers based on multiple linear regression. *Energy Reports*, 9, 150-162. <https://doi.org/10.1016/j.egyr.2023.01.020>
- Karantoumanis, E., & Ploskas, N. (2020). *Power Consumption Estimation in Data Centers Using Machine Learning Techniques* (pp. 195-200). https://doi.org/10.1007/978-3-030-53552-0_20
- Khan, T., Tian, W., Ilager, S., & Buyya, R. (2022). Workload forecasting and energy state estimation in cloud data centres: ML-centric approach. *Future Generation Computer Systems*, 128, 320-332. <https://doi.org/10.1016/j.future.2021.10.019>
- Kim, J. H., Shin, D. U., & Kim, H. (2024). Data Center Energy Evaluation Tool Development and Analysis of Power Usage Effectiveness with Different Economizer Types in Various Climate Zones. *Buildings*, 14(1), 299. <https://doi.org/10.3390/buildings14010299>
- Kumar, R., Khatri, S. K., & Diván, M. J. (2021). Power Usage Efficiency (PUE) Optimization with Counterpointing Machine Learning Techniques for Data Center Temperatures. *International Journal of Mathematical, Engineering and Management Sciences*, 6(6), 1594-1611. <https://doi.org/10.33889/IJMEMS.2021.6.6.095>
- Lascano, J., Saraguro, R., Chiza, L., Astudillo, J., & Guañuna, G. (2023). Estimación de la Demanda de una Estación de Carga para Vehículos Eléctricos Mediante la Aplicación de Métodos Probabilísticos y Predictivos. *Revista Técnica "energía"*, 20(1), 52-64. <https://doi.org/10.37116/revistaenergia.v20.n1.2023.569>
- Liao, X., Li, Y., Zhang, S., Wei, X., & Hu, J. (2026). Predicting GPU Training Energy Consumption in Data Centers Using Task Metadata via Symbolic Regression. *Energies*, 19(2), 448. <https://doi.org/10.3390/en19020448>
- Liu, J., Fan, R., Li, Z., Harnpornchai, N., & Qian, J. (2026). Proactive Cooling Control Algorithm for Data Centers Based on LSTM-Driven Predictive Thermal Analysis. *Applied System Innovation*, 9(1), 21. <https://doi.org/10.3390/asi9010021>
- Ma, H., & Ding, A. (2022). Method for evaluation on energy consumption of cloud computing data center based on deep reinforcement learning. *Electric Power Systems Research*, 208, 107899. <https://doi.org/10.1016/j.epsr.2022.107899>
- Misaghian, M. S. & others. (2023). Assessment of Carbon-Aware Flexibility Measures From Data Centres Using Machine Learning. *IEEE Transactions on Industry Applications*, 59(1), 70-81.
- Muraku, B., Haveriku, A., Deliu, A., & Mece, E. K. (2024). Data-Driven Prediction of Power Usage Effectiveness: A Machine Learning Case Study. *2024 32nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. <https://doi.org/10.23919/SoftCOM62040.2024.10721988>
- Ortega-Díaz, L. P., Cárdenas, J., & Osma-Pinto, G. (2023). Estrategias de predicción de consumo energético en edificaciones: Una revisión. *Tecnológicas*, 26(58), e2650.
- Rafique, A., Yu, X., Jawad, M., Song, Q., Yuan, Z., Sadiq, M. T., Daniel, K., & Shabbir, N. (2026). Two-Stage Optimization-Learning Framework for Uncertainty-Aware Multi-Zonal Data Center Energy Management. *Energies*, 19(7), 1736. <https://doi.org/10.3390/en19071736>
- Rantini, D., Najjiyya, S. I., Ghani, M., Yasmirullah, S. D. P., Fahmiyah, I., Ramadan, A., Othman, F., & Alya, N. A. (2026). Understanding the Spatial Distribution of Stunting in East Java, Indonesia: A Comparison of GWR and MS-GWR Models. *Statistics, Optimization and Information Computing*, 15(1), 311-323. <https://doi.org/10.19139/soic-2310-5070-3066>
- Reyes Almeida, F. J. (2021). *Implementación de un algoritmo para el seguimiento del punto de potencia máxima en un panel solar mediante un control predictivo por modelo*.
- Tarafdar, A., Sarkar, S., Das, R. K., & Khatua, S. (2023). Power Modeling for Energy-Efficient Resource Management in a Cloud Data Center. *Journal of Grid Computing*, 21(1), 10. <https://doi.org/10.1007/s10723-023-09642-5>

- Tenenaula Cunduri, M. R. (2025). *Desarrollo de un software simulador basado en modelos matemáticos y machine learning para analizar y optimizar redes híbridas de energía renovable* [Tesis de Posgrado, Universidad Nacional de Chimborazo (UNACH)]. <http://dspace.unach.edu.ec/handle/51000/16460>
- Teresa Riccio Barbosa, A., & Santiago Do Nascimento Lima, R. (2025). Estratégias de eficiência energética em data centers: Revisão sistemática. *Latin American Journal of Energy Research*, 12(4), 73-82. <https://doi.org/10.21712/lajer.2025.v12.n4.p73-82>
- Vergara, R. & others. (2025). Optimización del rendimiento de plantas solares en azotea mediante monitoreo IoT de bajo costo: Un estudio de caso en Montería, Colombia. *Revista de Ingeniería, Matemáticas y Ciencias de la Información*. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1909-83672025000100071
- Wang, X. & others. (2023). Energy Consumption Modeling in High-Performance Computing Racks. *Information Systems Research*. <https://doi.org/10.1287/isre.2023.1208>
- Wang, X., Zhou, S., Gong, Y., Liu, Y., & Liu, J. (2025). A Hierarchical Control Framework for HVAC Systems: Day-Ahead Scheduling and Real-Time Model Predictive Control Co-Optimization. *Energies*, 18(23), 6266. <https://doi.org/10.3390/en18236266>
- Yan, L., & Ming, T. (2023). Analysis on the Influence Factors of Heat Dissipation of Undersea Data Center Based on Multiple Linear Regression. *Procedia Comput. Sci.*, 228, 920-926. <https://doi.org/10.1016/j.procs.2023.11.121>
- Zapata, D., Meyer, M., & Müller, O. (2024). Combining Causal Discovery and Machine Learning for Modeling Data Center Operations. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2410.09516>
- Zhang, Y., & Liu, J. (2022). Prediction of Overall Energy Consumption of Data Centers in Different Locations. *Sensors*, 22(10), 3704. <https://doi.org/10.3390/s22103704>
- Zheng, Z. & others. (2022). Joinpoint–Multiple Linear Regression (JP–MLR) Model for Building Electricity Usage. *Energies*, 15(22), 8543. <https://doi.org/10.3390/en15228543>