ISSN: 1692-7257 - Volumen 1 – Número 47 - 2026

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Digital Object Identifier: 10.24054/rcta.v1i47.4356

Revisión sistemática sobre identificación de anomalías en el consumo de servicios web masivos mediante técnicas de machine learning

Systematic review on anomaly identification in massive web service consumption using machine learning techniques

Ing. Jesús Andrés Cruz Sanabria[1], Ing. Joaquín Iván Barrera Lozada[1], MSc. Karla Yohana Sánchez Mojica[1]

[1] ***Universidad de La Salle,*** *Facultad de Ingeniería, Maestría en Inteligencia Artificial, Bogotá, Cundinamarca, Colombia.*

Correspondencia: *{jcruz47, jbarrera17, kasanchez}@unisalle.edu.co*

Recibido: *20 septiembre 2025.* ***Aceptado:*** *18 diciembre 2025.* ***Publicado:*** *09 febrero 2026.*

Cómo citar: *J. A. Cruz Sanabria, J. I. Barrera Lozada y K. Y. Sánchez Mojica, "Revisión sistemática sobre identificación de anomalías en el consumo de servicios web masivos mediante técnicas de machine learning", RCTA, vol. 1, n.º. 47, pp. 209-218, feb. 2026. Recuperado de https://ojs.unipamplona.edu.co/index.php/rcta/article/view/4356*

Resumen: Este artículo presenta una revisión sistemática de la literatura sobre el uso de técnicas de Machine Learning aplicadas a la identificación de anomalías en servicios web y sistemas distribuidos. El proceso de revisión se desarrolló a partir de búsquedas estructuradas en bases de datos académicas reconocidas, incluyendo IEEE Xplore, Scopus, ScienceDirect y ACM Digital Library, considerando publicaciones entre 2021 y 2025. Se aplicaron criterios explícitos de inclusión y exclusión, lo que permitió seleccionar un conjunto final de cincuenta artículos relevantes. Los estudios analizados se organizaron según tipo de dato, enfoque de aprendizaje, dominio de aplicación y métricas empleadas, con el fin de identificar tendencias, fortalezas y limitaciones del estado del arte. Los resultados evidencian una creciente adopción de modelos híbridos y arquitecturas profundas, así como un interés sostenido por la explicabilidad y la escalabilidad en entornos distribuidos.

Palabras clave: revisión sistemática, detección de anomalías, machine learning, QoS, servicios web.

Abstract: This paper provides a systematic literature review on the application of machine learning techniques for anomaly detection within web services and distributed systems. Our methodology involved structured queries across major academic databases, including IEEE Xplore, Scopus, ScienceDirect, and the ACM Digital Library, covering research published between 2021 and 2025. Following the application of rigorous inclusion and exclusion criteria, a final cohort of 50 relevant articles was selected for detailed analysis. These studies were categorized based on data types, learning paradigms, application domains, and evaluation metrics to pinpoint current trends, strengths, and inherent limitations in the field. Our findings highlight a clear shift toward hybrid models and deep learning architectures, alongside a growing emphasis on explainability and scalability in distributed environments.

Keywords: literature review, anomaly detection, machine learning, QoS, web services.

1. INTRODUCCIÓN

Hoy en día, la infraestructura digital enfrenta una presión que no se ha visto anteriormente debido a la masificación del acceso a la red, que para el año 2025 ha alcanzado al 74% de la población mundial [1], [2], equivalente a 6 mil millones de individuos, tal como se ilustra en la Figura 1.

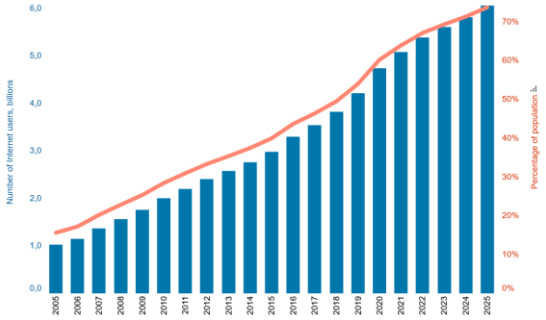

Fig. 1. *Número de usuario de internet*
Fuente: *Internation Telecommunication Union (ITU)*

Este crecimiento está estrechamente relacionado con la adopción de dispositivos móviles. A finales de 2024, aproximadamente 4400 millones de personas contaban con un teléfono inteligente para acceder a internet, lo que equivale al 54 % de la población mundial. Además, más del 80 % de estos accesos se realizan mediante redes 4G y 5G [3], complementado con el hecho de que, para inicios de 2025, se detectaron 5.56 mil millones de usuarios de internet, y que el 58% de los habitantes globales acceden mediante dispositivos propios. Lo anterior ha permitido que se consolide un ecosistema de servicios web masivos donde el tráfico de video representa el 75.9% del volumen total de datos celulares [1]. Este volumen de usuarios opera sobre una arquitectura de red de gran escala, en la cual el tráfico de datos ha crecido de forma exponencial. Se proyecta un consumo cercano a 1.5 zettabytes, lo que ha motivado la adopción de arquitecturas basadas en Machine Learning para la detección proactiva de anomalías [2] que garanticen el correcto consumo de servicios.

La infraestructura para esta demanda es de aproximadamente 6,111 centros de datos públicos a nivel mundial para finales de 2025 [4], cuya capacidad operativa enfrenta una presión técnica crítica debido a la integración de aplicaciones de Inteligencia Artificial (IA) generativa que requieren una potencia de procesamiento en la nube extrema comparada con otros servicios populares [4].

Aunque la cobertura de banda ancha móvil alcanza actualmente al 96% de la población global [2], la sofisticación de las redes 5G, que abarcan al 55.1% de la población mundial, agregan variables con una alta complejidad técnica que dificultan la monitorización y la detección de patrones de tráfico atípicos [2], [3].

En este contexto, la detección de anomalías adquiere un papel clave para preservar la estabilidad operativa de los proveedores de servicios web, especialmente en escenarios caracterizados por alta variabilidad en la demanda. El problema central reside en que los métodos tradicionales, basados en reglas estáticas y umbrales fijos, pueden ser insuficientes para administrar la naturaleza dinámica y los picos de demanda generados por eventos de alta concurrencia en plataformas de streaming, ecommerce, servicios financieros entre otros.

La complejidad técnica se ve presionada por la diversidad de los dispositivos y la transición hacia el 5G, tal como se mencionó anteriormente, que ya cubre al 55.1% de la población mundial, pero que coexiste con un 16% de usuarios que aún dependen de tecnologías 3G o dispositivos básicos [2], lo cual se puede observar en la figura 2, generando disparidades críticas en la Calidad de Servicio (QoS) y la Calidad de Experiencia (QoE).

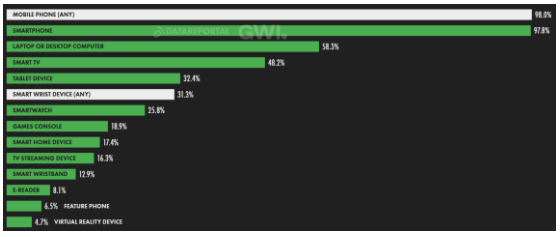

Fig. 2. *Tendencia de dispositivos usados para acceder a internet.* ***Fuente:*** *Digital 2025 Global Overview Report.*

El objetivo de este artículo consiste en realizar una revisión sistemática de literatura sobre métodos de identificación de anomalías en diferente campos web como servicios de streaming e IoT, entre otros, y establecer un marco de referencia de variables que puede servir para futuros estudios.

2. METODOLOGÍA

Este estudio se desarrolla como una revisión sistemática de literatura, buscando clasificar y sintetizar de manera estructurada el conocimiento existente sobre cómo se ha abordado la identificación de anomalías en el campo de Machine Learning. El enfoque de este es mixto, combinando el análisis cualitativo, examinando enfoques, tendencias y marcos conceptuales, con elementos cuantitativos, como la clasificación de artículos por tipo de datos, métodos de aprendizaje y objetivos.

Cabe aclarar que este estudio no tiene como objetivo la validación experimental de hipótesis, sino el análisis de cómo la comunidad científica ha abordado el problema de la detección de anomalías, qué soluciones se han propuesto y cuáles son los vacíos que aún persisten, ofreciendo una visión del estado del arte que sirva como base para investigaciones y desarrollos futuros.

2.2 Diseño metodológico

La revisión se centró en artículos científicos publicados entre los años 2021 y 2025, que abordaran la detección o predicción de anomalías utilizando técnicas de Machine Learning, aplicadas a servicios web, sistemas distribuidos, infraestructuras cloud, redes de telecomunicaciones e Internet de las Cosas (IoT). Considerando tanto investigaciones experimentales como revisiones y encuestas.

El proceso metodológico se apoyó en criterios de inclusión y exclusión previamente definidos, una estrategia de búsqueda estructurada y un método de análisis que permitió comparar los estudios de forma consistente.

2.3 Criterios de inclusión y exclusión

2.3.1 Criterios de inclusión

Se incluyeron en la revisión los artículos que cumplieran con los siguientes criterios:

- Publicaciones relacionadas con la detección o predicción de anomalías mediante Machine Learning.
- Estudios aplicados a servicios web, sistemas cloud, microservicios, redes, IoT o entornos digitales de gran escala.
- Revisiones de literatura o encuestas que analizaran tendencias, desafíos o comparaciones de técnicas.
- Artículos publicados entre 2021 y 2025.
- Artículos publicados en revistas cuyo cuartil sea 1 o 2. Se aceptan excepciones si el artículo está publicado en fuentes confiables.

2.3.2 Criterios de exclusión

Se excluyeron los estudios que presentaran alguna de las siguientes características:

- Artículos cuya fecha de publicación es anterior al 2021.
- Artículos con enfoques a otras industrias diferentes a las mencionadas en los criterios de inclusión.
- Artículos que no están relacionados con el uso de Machine Learning.
- Artículos publicados en revistas con quartiles inferiores al 2.

2.4 Muestra

En esta investigación, se encontraron 50 artículos científicos, compuestos por 60% correspondiente a estudios experimentales que proponen o evalúan modelos de Machine Learning para la detección de anomalías, y 40% correspondiente a revisiones de literatura y encuestas que resumen y analizan el estado del arte. Esto se representa en la Figura 3. Intencionalmente, se buscó mezclar evidencia práctica con análisis teóricos, facilitando una comprensión más completa del tema.

2.5 Instrumentos de recolección y análisis de información

Para poder decidir que recursos eran útiles para este estudio, y siguiendo los criterios de inclusión y exclusión, se creó una matriz de análisis, diseñada para recopilar de forma ordenada la información más relevante de cada artículo seleccionado. Esta herramienta permitió comparar los estudios de manera consistente y reducir posibles sesgos durante el análisis. La matriz incluyó los siguientes aspectos:

- Año de publicación
- Tipo de estudio
- Dominio de aplicación
- Tipo de datos utilizados
- Enfoque de aprendizaje (supervisado, no supervisado o mixto)
- Técnicas de Machine Learning empleadas
- Objetivo principal del estudio
- Principales resultados y limitaciones

Lo anterior se estableció basándonos en prácticas encontradas en revisiones sistemáticas previas, de las cuales algunas hacen parte de este estudio.

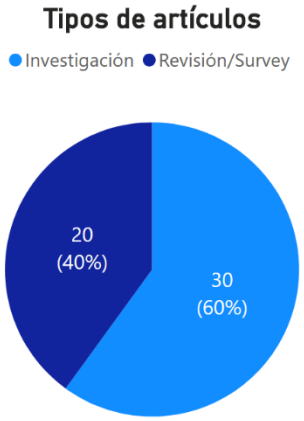

Fig. 3. *Tipos de artículos empleados.*
Fuente: *elaboración propia.*

2.6 Procedimiento

El desarrollo de la investigación se llevó a cabo en varias etapas consecutivas:

- Fase 1: Se delimitó el tema de investigación y se definieron los objetivos, centrados en analizar el uso de Machine Learning para la detección de anomalías en servicios web masivos.
- Fase 2: Se realizó una búsqueda sistemática en bases de datos académicas reconocidas, utilizando combinaciones de palabras clave relacionadas con detección de anomalías, Machine Learning, servicios web, cloud computing e IoT.
- Fase 3: Los artículos encontrados fueron evaluados inicialmente por título y resumen. Despues, se revisó el texto completo de los estudios preseleccionados para verificar el cumplimiento de los criterios que se establecieron.
- Fase 4: La información relevante de cada artículo se registró en la matriz de análisis. A partir de estos datos, se realizaron análisis descriptivos que permitieron identificar tendencias, enfoques predominantes y áreas en que se aplicaron.
- Fase 5: Los resultados se organizaron, construyendo la actual investigación y definiendo las bases para futuras líneas de trabajo.

3. RESULTADOS

Los resultados presentados en esta sección corresponden a una síntesis analítica de los hallazgos reportados en los estudios revisados, y no a experimentación propia realizada por los autores.

3.1. Evaluación cuantitativa de modelos y arquitecturas

El análisis de los estudios que reportan validaciones cuantitativas, resumido en la Tabla 1, sugiere que la especialización por dominio es un factor recurrente asociado al desempeño de los modelos evaluados [5], [6], [7]. En los estudios revisados sobre entornos de alta velocidad, como IoT y redes 6G, se reporta un desempeño destacado de técnicas basadas en LSH y enfoques híbridos, con precisiones cercanas al 100 % en determinados conjuntos de datos [8], [9]. Por el contrario, en dominios de seguridad crítica o medicina, las arquitecturas de grafos y atención (GNN, Transformers) dominan la captura de dependencias no lineales [10], [11].

Tabla 1: *Análisis sobre evaluación, propuesta o comparación de modelos con métricas de rendimiento.*

Modelo / Algoritmo	Dataset de Referencia	Métrica Destacada
LSH + Random Forest [8]	ToN-IoT, MQTT-IoT	Exactitud: 99.82%
HABBAs (AdaBoost + Bagging) [12]	CICDDOS2019	Exactitud: 99.95%
AE (Vector Reconstruction Error) [13]	CIDDS-001	F1-score: 100%
WDLog (Wide & Deep Learning) [14]	HDFS, BGL	F1-score > 90%
CCTAK (TCN + KAN + VAE) [15]	SKAB, SWaT	AUC-ROC: 0.8191
GIN (Graph Att. + Informer) [10]	MSL, SMAP	F1-score: 0.9604
SMOTETomek + ML Models [16]	WSN-DS	Exactitud: 99.92%
Diner (Memory AE + SNR) [17]	GAIA, NAB	F1-score: 0.706
AnyLog (BERT + SOM + AE) [18]	HDFS, BGL	Exactitud: 95.0%
AADS (Online Clustering) [19]	Breast Cancer, Ionosphere	F1-score: 76.13%
DGMM + ML Ensemble [20]	Tráfico Celular	RMSE: 0.026
XMLAD (Decision Tree Logic) [21]	NokiaFL (Real)	Recall: 100%
MDI vs Deep Learning [22]	UCR Archive	AUC-ROC: 0.66
Deep Isolation Forest (DIF) [23], [24]	Tabular / Graph / TS	AUC-PR: +144% vs iF

C-LSTM-AE [25]	Yahoo Webscope S5	Mejor F1 vs CNN/RNN
LSTM-Markov Hybrid [26]	Smart City Sensors	Eficiencia AD: 96.03%
HGN2HIA (Heterog. Graph Att.) [27]	QoS Web Services	p-value: 0.005
DDQN-PER (RL) [28]	Occupancy Detection	Recall: 97.10%
TTrees (CIAN Methodology) [29]	MONROE, Nokia	Exactitud: 99.6%
Generalized iForest (GIF) [30]	Benchmarks (Aloi)	Superior AUC-PR
CNN (1D, 2D, 3D) Models [6]	IoT-DS-2	Accuracy: >99.7%
Multi-method TS Evaluation [31]	SWaT, SMD	LSTM AUC: 0.863
Feedback K-means [32]	Web Service QoS	RMSE: 0.051
CAWAL Framework [7]	Web Portal Logs	Exactitud: 92.5%
XGBoost (African J.) [33]	CICIDS2017	F1-score: 0.987
VAE Stability Analysis [34]	Wireless Comm.	134 anomalías det.
POT Threshold Selection [35]	Firewall logs	70% of Opt. MCC

Fuente: *elaboración propia.*

3.2. Síntesis teórica y taxonómica de la literatura

En la tabla 2 se observa un análisis sobre aquella literatura cuyos resultados se enfocan en identificar hallazgos o desafíos relevantes, tal como la evolución desde el filtrado estadístico métodos de detección de anomalías explicable (XAD) [36], [37].

Algunos trabajos revisados señalan el fenómeno conocido como “Clever Hans”, en el cual modelos complejos alcanzan altas precisiones apoyándose en correlaciones espurias del conjunto de datos [36].

Tabla 2: *Análisis sobre artículos de revisión, taxonomía y marcos teóricos.*

Foco	Datasets/Benchmarks Analizados	Hallazgo o Desafío Destacado
Taxonomía de 52 algoritmos [38], [39]	KDD Cup 99/98, NSL-KDD, UCI Repository	Clasificación en 7 mecanismos; los métodos de aislamiento son los más escalables.
XAD (Detección Explicable) [36], [40]	Enron, Yelp, Amazon, Twitter Sybil, Elliptic	Identificación del efecto "Clever Hans": precisión basada en ruido en lugar de causalidad.
Deep Transfer Learning (DTL) [41]	CWRU, IMS, PHM 2012 (Series Industriales)	Riesgo de transferencia negativa si la brecha de dominio es excesiva.
Graph Anomaly Detection (GAD) [40]	DBLP, Wikipedia, Reddit, Amazon, Enron	Sistematización de anomalías estructurales en nodos, aristas y subgrafos.
GNN en entornos IIoT [11]	SWaT, WADI, BATADAL, Xcos, epanetCPA	Necesidad de modelar la interdependencia relacional evolutiva en sistemas ciberfísicos.
Knowledge-based Systems (KBS) [37]	NSL-KDD, UNSW-NB15, DS2OS	Los sistemas semánticos ofrecen mayor interpretabilidad, pero enfrentan retos de actualización.
DL para Detección en Logs [42]	HDFS, BGL, Thunderbird, Spirit, OpenStack	Desafío de la inestabilidad de plantillas y el volumen masivo de datos no estructurados.
MTSAD (Series Multivariantes) [43]	CHB-MIT (EEG), Gas Pipeline, Yahoo Webscope	Clasificación por granularidad: anomalías de punto, de intervalo y de serie completa.
Tráfico de Red Cifrado (SSL/TLS) [44]	CTU-13, CIC-IDS-2017, UNSW-NB15, MTA	Inviabilidad de inspección profunda (DPI); dependencia de características estadísticas de flujo.
AutoML en Detección de Anomalías [45]	Benchmarks generales de Outlier Detection	El problema CASH: dificultad de automatizar la selección y ajuste sin etiquetas de verdad fundamental.
AD en Smart Environments [46]	Yahoo! S5, NAB, UCR Archive, Space Shuttle	Retos de escasez de etiquetas y contextualización en entornos inteligentes.
Microservicios y RCA [47]	KPIs, trazas y registros de "Sock Shop"	El análisis de causa raíz (RCA) requiere correlación entre

		KPIs y logs de servicios.
Federated Machine Learning [48], [49]	Datasets distribuidos de red y dispositivos móviles	Balance entre privacidad de datos y eficiencia estadística en el entrenamiento descentralizado.
Federated Learning para IoT [48], [49]	CIFAR-10, MNIST, Imagenet (usados como proxies)	La divergencia de pesos en datos no-IID reduce la precisión hasta un 55%.
SLR de Machine Learning (General) [39]	KDD Cup 99, NSL-KDD, UCI, Real-life datasets	Solo el 27% de las investigaciones registradas utilizan métodos puramente no supervisados.
Sistemas de Sensores (Multi-pers.) [50]	Bot-IoT, ODDS, NAB, Yahoo Webscope, ELKI	Importancia de la hibridación estadística y profunda para streaming de datos.
Streaming de Multimedia (User-centric) [51]	LIVE Netflix, CSIQ, LFOVIA, MCQoE, FCC	Modelado de la Calidad de Experiencia (QoE) mediante métricas subjetivas y objetivas.
AD para Fallos en Redes WCN [34]	GuifiSants (Producción real), Linux Kernel features	La inclusión de métricas de hardware (CPU/RAM) aumenta la capacidad de detección de fallos de red.
LSTM en Sistemas Técnicos [23], [52]	Amazon, Wireless Sensor Network, Electricity consumption	Las arquitecturas encoder-decoder son superiores para aprender relaciones temporales estacionarias y no estacionarias.
Time-Series DL (Guidelines)	SWaT, SMAP, WADI, SMD, MSL	El ruido es el factor crítico que complica la detección en sistemas de control industrial (CPS).

Fuente: *elaboración propia.*

3.3. Análisis de operatividad en tiempo real

El tercer eje de este análisis es acerca de la transición crítica de los modelos de detección desde entornos de laboratorio hacia infraestructuras de producción reales, donde la sostenibilidad operativa y la exactitud temporal son requisitos obligatorios.

En este contexto, la investigación de [9] resulta disruptiva al proponer el algoritmo StreamWNNov, el cual introduce una distinción semántica vital entre el concepto de "novedad", entendido como un patrón emergente que debe enseñarse modelo, y el de "anomalía", que es una excepción que dispara alertas inmediatas.

Los resultados obtenidos sobre la demanda eléctrica española validan que este enfoque de aprendizaje incremental permite reducir el error de predicción (MAPE) al 2.07% en flujos continuos, manteniendo una complejidad computacional de O(1) en su fase online [53]. Esta capacidad de autoactualización garantiza la viabilidad del procesamiento en tiempo real sin incurrir en los costos prohibitivos de un reentrenamiento masivo, un factor de escalabilidad que resuena con las advertencias de [53]. Según el análisis de tendencias de estos últimos, existe una brecha crítica en el estado del arte actual: la mayoría de los modelos de detección de intrusiones (IDS) para redes IoT se evalúan de forma estática u offline, ignorando que la superficie de ataque masiva y la carga computacional del Deep Learning pueden invalidar la respuesta inmediata necesaria ante amenazas de "día cero" en el borde (edge) de la red.

Por último, la convergencia técnica entre la inteligencia artificial y los sistemas de tiempo real (RTS) es examinada por [54], quien subraya que en dominios de seguridad crítica, como drones o robótica industrial, el cumplimiento del tiempo de ejecución del peor caso (WCET) y el respeto estricto a los tiempos es tan vital como la propia precisión de los algoritmos. Estos concluyen que el diseño de sistemas de alto impacto debe basarse en un equilibrio sistémico donde la sofisticación de la arquitectura se armonice con las restricciones de hardware heterogéneo y la latencia determinista exigida por el entorno físico supervisado.

3.4 Discusión de Tendencias y Resultados

El análisis de los estudios revisados permite identificar varios patrones recurrentes, aunque no todos se manifiestan con la misma fuerza en cada dominio. A continuación, se discuten aquellos que aparecen con mayor frecuencia en la literatura reciente:

- Aunque múltiples estudios reportan un mejor desempeño del Deep Learning en escenarios multivariantes de gran escala, este dominio no es absoluto. Algunos trabajos muestran que, bajo

restricciones de datos o latencia, enfoques más simples pueden ofrecer resultados comparables.

- Transición histórica, desde el 2021 al 2025, en donde se puede identificar un desplazamiento desde métodos de proximidad (k-NN, SVM) enfocados en anomalías de punto, hacia arquitecturas de grafos dinámicos que modelan la interdependencia de sensores en redes 6G.
- Varios autores coinciden en que la precisión, considerada de forma aislada, resulta insuficiente para evaluar el desempeño real de los sistemas de detección de anomalías. La validación científica actual exige técnicas post-hoc como SHAP para desglosar la contribución de cada sensor, transformando la "caja negra" en una herramienta de diagnóstico.
- Los modelos basados en aislamiento (Isolation Forest) presentan la mayor escalabilidad, siendo los únicos viables para despliegues de tiempo real en dispositivos de borde con energía limitada.
- Los resultados reportados para algoritmos como StreamWNNov sugieren que la capacidad de distinguir entre “novedad” y “anomalía” es un aspecto relevante para la sostenibilidad de sistemas.

4. CONCLUSIONES

La detección de anomalías ha evolucionado, pasando de ser una tarea estadística periférica para convertirse en el pilar de la resiliencia en ecosistemas digitales que están hiperconectados. Tras evaluar 50 artículos recientes, se concluye que no existe una solución universal; la efectividad de un sistema depende de la armonía entre la naturaleza del dato, los requisitos de latencia y la necesidad de transparencia en la toma de decisiones. Mientras que el aprendizaje profundo domina en la identificación de irregularidades en flujos multivariantes complejos, los métodos clásicos y los basados en conocimiento conservan una relevancia indiscutible debido a su eficiencia computacional e interpretabilidad nativa.

Diversos autores sugieren que la incorporación de explicabilidad, aprendizaje federado y mecanismos autónomos podría convertirse en una línea dominante de investigación en los próximos años, especialmente en entornos regulados o distribuidos.

La capacidad de proporcionar un razonamiento tangible detrás de cada alerta no solo aumenta la confianza del usuario, sino que facilita la prevención proactiva de fallos sistémicos. Por último, la integración de marcos de aprendizaje incremental permitirá que la IA se adapte dinámicamente a un mundo volátil, transformando simples alertas en conocimiento accionable para la estabilidad de las infraestructuras críticas globales

REFERENCIAS

[1] Meltwater, “Digital 2025 Global Overview Report,” Meltwater, 1, Feb. 2025. [Online]. Available: https://drive.google.com/file/d/1fonmdNLLMbVFB9f-azux0ZTJPsw1CeZ8/view

[2] “Measuring digital development: Facts and Figures 2025,” ITU. Accessed: Feb. 04, 2026. [Online]. Available: https://www.itu.int/hub/publication/d-ind-ict_mdd-2025-3/

[3] “The State of Mobile Internet Connectivity 2025: Overview Report | GSMA Intelligence.” Accessed: Feb. 05, 2026. [Online]. Available: https://www.gsmaintelligence.com/research/the-state-of-mobile-internet-connectivity-2025-overview-report

[4] “How Many Data Centers Are There and Where Are They Being Built?” Accessed: Feb. 05, 2026. [Online]. Available: https://www.abiresearch.com/blog/data-centers-by-region-size-company

[5] K. Choi, J. Yi, C. Park, and S. Yoon, “Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines,” *IEEE Access*, vol. 9, pp. 120043–120065, 2021, doi: 10.1109/ACCESS.2021.3107975.

[6] I. Ullah and Q. H. Mahmoud, “Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks,” *IEEE Access*, vol. 9, pp. 103906–103926, 2021, doi: 10.1109/ACCESS.2021.3094024.

[7] Ö. Canay and Ü. Kocabıçak, “Predictive modeling and anomaly detection in large-scale web portals through the CAWAL framework,” *Knowledge-Based Systems*, vol. 306, p. 112710, Dec. 2024, doi: 10.1016/j.knosys.2024.112710.

[8] M. L. Hernandez-Jaimes, A. Martinez-Cruz, and K. A. Ramírez-Gutiérrez, “A Machine Learning approach for anomaly detection on the Internet of Things based on Locality-Sensitive Hashing,” *Integration*, vol. 96, p. 102159, May 2024, doi: 10.1016/j.vlsi.2024.102159.

[9] S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, “Machine Learning and Deep

Learning Techniques for Internet of Things Network Anomaly Detection—Current Research Trends," *Sensors*, vol. 24, no. 6, 2024, doi: 10.3390/s24061968.

[10] C. Wang and G. Liu, "From anomaly detection to classification with graph attention and transformer for multivariate time series," *Advanced Engineering Informatics*, vol. 60, p. 102357, Apr. 2024, doi: 10.1016/j.aei.2024.102357.

[11] Y. Wu, H. -N. Dai, and H. Tang, "Graph Neural Networks for Anomaly Detection in Industrial Internet of Things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214–9231, Jun. 2022, doi: 10.1109/JIOT.2021.3094295.

[12] M. M. Saeed, R. A. Saeed, M. Abdelhaq, R. Alsaqour, M. K. Hasan, and R. A. Mokhtar, "Anomaly Detection in 6G Networks Using Machine Learning Methods," *Electronics*, vol. 12, no. 15, 2023, doi: 10.3390/electronics12153300.

[13] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, p. 1, Jan. 2023, doi: 10.1186/s42400-022-00134-9.

[14] W. Niu, X. Liao, S. Huang, Y. Li, X. Zhang, and B. Li, "A robust Wide & Deep learning framework for log-based anomaly detection," *Applied Soft Computing*, vol. 153, 2024, doi: 10.1016/j.asoc.2024.111314.

[15] Y. Abudurexiti, G. Han, F. Zhang, and L. Liu, "An explainable unsupervised anomaly detection framework for Industrial Internet of Things," *Computers & Security*, vol. 148, p. 104130, Jan. 2025, doi: 10.1016/j.cose.2024.104130.

[16] Md. A. Talukder, S. Sharmin, M. A. Uddin, M. M. Islam, and S. Aryal, "MLSTL-WSN: machine learning-based intrusion detection using SMOTETomek in WSNs," *Int. J. Inf. Secur.*, vol. 23, no. 3, pp. 2139–2158, Jun. 2024, doi: 10.1007/s10207-024-00833-z.

[17] Y. Jing *et al.*, "Diner: Interpretable Anomaly Detection for Seasonal Time Series in Web Services," *IEEE Transactions on Services Computing*, vol. 17, no. 5, pp. 2248–2260, Oct. 2024, doi: 10.1109/TSC.2024.3422894.

[18] A. Aziz and K. Munir, "Anomaly Detection in Logs Using Deep Learning," *IEEE Access*, vol. 12, pp. 176124–176135, 2024, doi: 10.1109/ACCESS.2024.3506332.

[19] M. Y. Iqbal Basheer *et al.*, "Autonomous anomaly detection for streaming data," *Knowledge-Based Systems*, vol. 284, p. 111235, Jan. 2024, doi: 10.1016/j.knosys.2023.111235.

[20] A. Gorshenin, A. Kozlovskaya, S. Gorbunov, and I. Kochetkova, "Mobile network traffic analysis based on probability-informed machine learning approach," *Computer Networks*, vol. 247, p. 110433, Jun. 2024, doi: 10.1016/j.comnet.2024.110433.

[21] J. M. Ramírez, F. Díez, P. Rojo, V. Mancuso, and A. Fernández-Anta, "Explainable machine learning for performance anomaly detection and classification in mobile networks," *Computer Communications*, vol. 200, pp. 113–131, Feb. 2023, doi: 10.1016/j.comcom.2023.01.003.

[22] F. Rewicki, J. Denzler, and J. Niebling, "Is It Worth It? Comparing Six Deep and Classical Methods for Unsupervised Anomaly Detection in Time Series," *Applied Sciences*, vol. 13, no. 3, 2023, doi: 10.3390/app13031778.

[23] U. A. Usmani, I. Abdul Aziz, J. Jaafar, and J. Watada, "Deep Learning for Anomaly Detection in Time-Series Data: An Analysis of Techniques, Review of Applications, and Guidelines for Future Research," *IEEE Access*, vol. 12, pp. 174564–174590, 2024, doi: 10.1109/ACCESS.2024.3495819.

[24] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep Isolation Forest for Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12591–12604, Dec. 2023, doi: 10.1109/TKDE.2023.3270293.

[25] C. Yin, S. Zhang, J. Wang, and N. N. Xiong, "Anomaly Detection Based on Convolutional Recurrent Autoencoder for IoT Time Series," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 1, pp. 112–122, Jan. 2022, doi: 10.1109/TSMC.2020.2968516.

[26] S. V. and S. A., "LSTM-Markov based efficient anomaly detection algorithm for IoT environment," *Applied Soft Computing*, vol. 136, p. 110054, Mar. 2023, doi: 10.1016/j.asoc.2023.110054.

[27] S. Lv, F. Yi, P. He, and C. Zeng, "QoS Prediction of Web Services Based on a Two-Level Heterogeneous Graph Attention Network," *IEEE Access*, vol. 10, pp. 1871–1880, 2022, doi: 10.1109/ACCESS.2021.3138127.

[28] D. Fährmann, N. Jorek, N. Damer, F. Kirchbuchner, and A. Kuijper, "Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments," *IEEE Access*, vol. 10, pp.

60836–60848, 2022, doi: 10.1109/ACCESS.2022.3179720.

[29] M. Moulay, R. G. Leiva, P. J. Rojo Maroni, F. Diez, V. Mancuso, and A. Fernández Anta, "Automated identification of network anomalies and their causes with interpretable machine learning: The CIAN methodology and TTrees implementation," *Computer Communications*, vol. 191, pp. 327–348, Jul. 2022, doi: 10.1016/j.comcom.2022.05.013.

[30] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tourneret, "Generalized isolation forest for anomaly detection," *Pattern Recognition Letters*, vol. 149, pp. 109–119, Sep. 2021, doi: 10.1016/j.patrec.2021.05.022.

[31] M. A. Belay, S. S. Blakseth, A. Rasheed, and P. Salvo Rossi, "Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions," *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052844.

[32] Y. Song, "Web service reliability prediction based on machine learning," *Computer Standards & Interfaces*, vol. 73, p. 103466, Jan. 2021, doi: 10.1016/j.csi.2020.103466.

[33] V. R. Gudelli, "Anomaly Detection in Cloud Networks Using Machine Learning Algorithms," Jun. 2024, doi: 10.5281/ZENODO.15271016.

[34] L. Cerdà-Alabern, G. Iuhasz, and G. Gemmi, "Anomaly detection for fault detection in wireless community networks using machine learning," *Computer Communications*, vol. 202, pp. 191–203, Mar. 2023, doi: 10.1016/j.comcom.2023.02.019.

[35] A. Komadina, M. Martinić, S. Groš, and Ž. Mihajlović, "Comparing Threshold Selection Methods for Network Anomaly Detection," *IEEE Access*, vol. 12, pp. 124943–124973, 2024, doi: 10.1109/ACCESS.2024.3452168.

[36] Z. Li, Y. Zhu, and M. Van Leeuwen, "A Survey on Explainable Anomaly Detection," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 1, pp. 1–54, Jan. 2024, doi: 10.1145/3609333.

[37] A. Q. Khan, S. El Jaouhari, N. Tamani, and L. Mroueh, "Knowledge-based anomaly detection: Survey, challenges, and future directions," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108996, Oct. 2024, doi: 10.1016/j.engappai.2024.108996.

[38] D. Samariya and A. Thakkar, "A Comprehensive Survey of Anomaly Detection Algorithms," *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, Jun. 2023, doi: 10.1007/s40745-021-00362-9.

[39] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.

[40] X. Ma *et al.*, "A Comprehensive Survey on Graph Anomaly Detection With Deep Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12012–12038, Dec. 2023, doi: 10.1109/TKDE.2021.3118815.

[41] P. Yan *et al.*, "A Comprehensive Survey of Deep Transfer Learning for Anomaly Detection in Industrial Time Series: Methods, Applications, and Directions," *IEEE Access*, vol. 12, pp. 3768–3789, 2024, doi: 10.1109/ACCESS.2023.3349132.

[42] M. Landauer, S. Onder, F. Skopik, and M. Wurzenberger, "Deep learning for anomaly detection in log data: A survey," *Machine Learning with Applications*, vol. 12, p. 100470, Jun. 2023, doi: 10.1016/j.mlwa.2023.100470.

[43] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion*, vol. 91, pp. 93–102, Mar. 2023, doi: 10.1016/j.inffus.2022.10.008.

[44] I. H. Ji, J. H. Lee, M. J. Kang, W. J. Park, S. H. Jeon, and J. T. Seo, "Artificial Intelligence-Based Anomaly Detection Technology over Encrypted Traffic: A Systematic Literature Review," *Sensors*, vol. 24, no. 3, p. 898, Jan. 2024, doi: 10.3390/s24030898.

[45] M. Bahri, F. Salutari, A. Putina, and M. Sozio, "AutoML: state of the art with a focus on anomaly detection, challenges, and research directions," *International Journal of Data Science and Analytics*, vol. 14, no. 2, pp. 113–126, Aug. 2022, doi: 10.1007/s41060-022-00309-0.

[46] D. Fährmann, L. Martín, L. Sánchez, and N. Damer, "Anomaly Detection in Smart Environments: A Comprehensive Survey," *IEEE Access*, vol. 12, pp. 64006–64049, 2024, doi: 10.1109/ACCESS.2024.3395051.

[47] J. Soldani and A. Brogi, "Anomaly Detection and Failure Root Cause Analysis in (Micro) Service-Based Cloud Applications: A Survey," *ACM Comput. Surv.*, vol. 55, no. 3, pp. 1–39, Mar. 2023, doi: 10.1145/3501297.

[48] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated Learning for Internet of Things: A Comprehensive Survey," *IEEE Communications Surveys & Tutorials*, vol. 23,

no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.

[49] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated Machine Learning: Survey, Multi-Level Classification, Desirable Criteria and Future Directions in Communication and Networking Systems," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1342–1397, 2021, doi: 10.1109/COMST.2021.3058573.

[50] L. Erhan *et al.*, "Smart anomaly detection in sensor systems: A multi-perspective review," *Information Fusion*, vol. 67, pp. 64–79, Mar. 2021, doi: 10.1016/j.inffus.2020.10.001.

[51] M. Ghosh and C. Singhal, "A review on machine learning based user-centric multimedia streaming techniques," *Computer Communications*, vol. 231, p. 108011, Feb. 2025, doi: 10.1016/j.comcom.2024.108011.

[52] B. Lindemann, B. Maschler, N. Sahlab, and M. Weyrich, "A survey on anomaly detection for technical systems using LSTM networks," *Computers in Industry*, vol. 131, p. 103498, Oct. 2021, doi: 10.1016/j.compind.2021.103498.

[53] L. Melgar-García, D. Gutiérrez-Avilés, C. Rubio-Escudero, and A. Troncoso, "Identifying novelties and anomalies for incremental learning in streaming time series forecasting," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106326, Aug. 2023, doi: 10.1016/j.engappai.2023.106326.

[54] J. Bian *et al.*, "Machine Learning in Real-Time Internet of Things (IoT) Systems: A Survey," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8364–8386, Jun. 2022, doi: 10.1109/JIOT.2022.3161050.