DETECCIÓN DE ANOMALÍAS EN SERVICIOS CLOUD UTILIZANDO TÉCNICAS DE APRENDIZAJE NO SUPERVISADO

REVISTA UGC | Revista científica de la Universidad del Golfo de California
Volumen 3 | Número 3 | Septiembre - 2025

DETECTION OF ANOMALIES IN CLOUD SERVICES USING UNSUPERVISED LEARNING TECHNIQUES

Diana Carolina Decimavilla-Alarcón[1]
E-mail: ddecimavilla@istvr.edu.ec
ORCID: https://orcid.org/0000-0002-0375-0216
Carlos Andrés Wilson-Alvarado[1]
E-mail: ca.wilson@istvr.edu.ec
ORCID: https://orcid.org/0009-0004-1602-0281
[1] Instituto Superior Tecnológico Vicente Rocafuerte. Ecuador.

Cita sugerida (APA, séptima edición)

Decimavilla-Alarcón, D. C., & Wilson-Alvarado, C. A. (2025). Detección de anomalías en servicios Cloud utilizando técnicas de aprendizaje no supervisado. *Revista UGC, 3(3), 175-185.*

Fecha de presentación: 22/05/2025 **Fecha de aceptación:** 13/07/2025 **Fecha de publicación:** 01/09/2025

RESUMEN

La detección temprana de anomalías en servicios cloud se ha convertido en un desafío crítico debido a la creciente complejidad de las infraestructuras tecnológicas. Este estudio tiene como objetivo analizar y evaluar la efectividad de las técnicas de aprendizaje no supervisado para la detección de anomalías en entornos cloud computing, proporcionando un marco analítico comprehensivo que facilite la selección e implementación de algoritmos apropiados según contextos específicos. La investigación emplea una metodología cualitativa con enfoque descriptivo bibliográfico, fundamentada en el análisis sistemático de literatura científica especializada publicada entre 2018 y 2024, incluyendo el examen detallado de casos de estudio que demuestran la aplicación práctica de estas técnicas en entornos reales. Los hallazgos revelan que la efectividad de los algoritmos varía significativamente según el contexto de aplicación, donde técnicas como Isolation Forest destacan en el manejo de datos de alta dimensionalidad, alcanzando niveles de precisión del 99% en la detección de fraudes cuando se implementan en conjunto con herramientas de procesamiento distribuido. Asimismo, se evidencia una tendencia hacia la implementación de soluciones híbridas que combinan múltiples técnicas, como demuestra el sistema SAGE al integrar Redes Bayesianas Causales con Autoencoders Variacionales para la identificación efectiva de problemas de rendimiento en sistemas de microservicios. Finalmente, se establece que la detección efectiva de anomalías en servicios cloud requiere un enfoque adaptativo que combine múltiples técnicas de aprendizaje no supervisado, cuya selección debe basarse en las características específicas del entorno y los requisitos particulares de cada implementación. Esta investigación no solo contribuye al avance metodológico en el campo, sino que también proporciona lineamientos prácticos para optimizar la seguridad y eficiencia de las infraestructuras cloud computing.

Palabras clave:

Aprendizaje no supervisado, detección de anomalías, computación en la nube.

ABSTRACT

Early anomaly detection in cloud services has become a critical challenge due to the increasing complexity of technological infrastructures. This study aims to analyze and evaluate the effectiveness of unsupervised learning techniques for anomaly detection in cloud computing environments, providing a comprehensive analytical framework to facilitate the selection and implementation of appropriate algorithms according to specific contexts. The research employs a qualitative methodology with a descriptive bibliographic approach, based on the systematic analysis of specialized scientific literature published between 2018 and 2024, including a detailed examination of case studies that demonstrate the practical application of these techniques in real-world environments. The findings reveal that the effectiveness of algorithms varies significantly according to the application context, where techniques such as Isolation Forest excel in handling high-dimensional data, achieving accuracy levels of

99% in fraud detection when implemented in conjunction with distributed processing tools. Likewise, there is evidence of a trend towards the implementation of hybrid solutions that combine multiple techniques, as demonstrated by the SAGE system, which integrates Causal Bayesian Networks with Variational Autoencoders for the effective identification of performance problems in microservices systems. Finally, it is established that effective anomaly detection in cloud services requires an adaptive approach that combines multiple unsupervised learning techniques, the selection of which should be based on the specific characteristics of the environment and the particular requirements of each implementation. This research not only contributes to methodological advancement in the field but also provides practical guidelines for optimizing the security and efficiency of cloud computing infrastructures.

Keywords:

Unsupervised learning, anomaly detection, cloud computing.

INTRODUCCIÓN

Las técnicas de aprendizaje no supervisado han experimentado una transformación exponencial en el panorama tecnológico a nivel mundial, llegando a considerarse como una herramienta fundamental para la comprensión de sistemas complejos en múltiples disciplinas científicas, a escala internacional (Acosta-Servín et al., 2025; Casimiro- et al., 2025; Lavado-Rojas et al., 2025) estas metodologías han revolucionado la capacidad de análisis de datos en campos que comprenden desde la genómica hasta la ciberseguridad, permitiendo la identificación de patrones intrínsecos sin requerir etiquetado previo de información (Priarone et al.,, 2024), por otro lado, el aprendizaje no supervisado representa una aproximación metodológica que facilita la extracción de conocimiento en entornos caracterizados por alta incertidumbre y complejidad computacional, transformando paradigmas tradicionales de procesamiento de datos (Li, 2023).

El desarrollo de las tecnologías, actualmente ha potenciado la relevancia de estas técnicas, particularmente en infraestructuras cloud computing donde la gestión de grandes volúmenes de datos demanda aproximaciones innovadoras para comprender comportamientos dinámicos, en este sentido, el crecimiento exponencial de servicios cloud, genera ecosistemas tecnológicos cada vez más complejos y heterogéneos que requieren metodologías avanzadas de monitoreo y detección de anomalías (Saikiran et al., 2024), razón por la cual, la comunidad científica internacional reconoce que la implementación efectiva de técnicas de aprendizaje no supervisado puede optimizar significativamente la eficiencia, seguridad y resiliencia de infraestructuras computacionales distribuidas.

En el contexto latinoamericano, la adopción de servicios cloud ha experimentado un crecimiento significativo, generando desafíos únicos en términos de seguridad y gestión de infraestructuras tecnológicas, países como Brasil, Argentina y Colombia han incrementado su inversión en infraestructura digital, demandando soluciones especializadas para detección de anomalías que consideren las particularidades regionales. Además, según Jurado et al. (2020), es fundamental desarrollar marcos analíticos que tenga en cuenta las limitaciones de infraestructura y los desafíos específicos de escalabilidad que enfrentan las microempresas en Latinoamérica, estas consideraciones son esenciales para implementar tecnologías de seguridad de la información que respondan a las necesidades particulares de este sector.

La problemática central de la investigación se enfoca en la limitación de las metodologías convencionales para detectar anomalías en servicios cloud computing, particularmente en su capacidad para adaptarse rápidamente a patrones emergentes y contextos dinámicos, problemática que se acentúa considerando la naturaleza distribuida y altamente escalable de los servicios cloud, donde los métodos estadísticos clásicos pierden su efectividad ante la complejidad de los datos y la variabilidad de los comportamientos computacionales. La investigación busca generar una evaluación comprehensiva de técnicas de aprendizaje no supervisado, proporcionando lineamientos científicos para optimizar la detección de comportamientos atípicos en infraestructuras tecnológicas, el objetivo fundamental radica en comprender las capacidades, restricciones y potencial aplicabilidad de algoritmos no supervisados en la identificación temprana de eventos críticos en servicios cloud.

El objetivo general de este estudio es analizar técnicas de aprendizaje no supervisado para la detección efectiva de anomalías en servicios cloud computing.

METODOLOGÍA

La presente investigación adopta un diseño cualitativo con enfoque descriptivo bibliográfico, fundamentado en el análisis exhaustivo de literatura científica especializada en técnicas de aprendizaje no supervisado para la detección de anomalías en servicios cloud. El método principal empleado es la revisión documental sistemática, que ha permitido examinar y sintetizar un corpus significativo de fuentes académicas, incluyendo artículos científicos, libros especializados y casos de estudio publicados entre 2018 y 2024. Esta aproximación metodológica resulta particularmente apropiada dado que el objetivo principal de la investigación es desarrollar una comprensión profunda de las capacidades y limitaciones de los algoritmos de aprendizaje no supervisado en contextos cloud, para lo cual se ha realizado un análisis interpretativo de las fuentes, estableciendo conexiones conceptuales y

evaluando la aplicabilidad práctica de las diferentes técnicas estudiadas. La investigación se estructura en dos fases principales: primero, una exploración comprehensiva de las bases teóricas y fundamentos metodológicos del aprendizaje no supervisado en detección de anomalías; y segundo, un análisis detallado de casos de estudio que ejemplifican la implementación práctica de estas técnicas en entornos cloud reales.

El proceso de análisis de datos se ha realizado mediante una aproximación cualitativa interpretativa, empleando técnicas de análisis de contenido para identificar patrones, tendencias y relaciones significativas en la literatura examinada. Los resultados se han organizado y presentado mediante tablas comparativas que sintetizan las características, ventajas y desafíos de los diferentes algoritmos estudiados; así como mediante el análisis detallado de casos de estudio que proporcionan evidencia empírica sobre la efectividad de estas técnicas en implementaciones reales. Esta metodología ha permitido no solo catalogar y describir las diferentes aproximaciones técnicas disponibles, sino también desarrollar un marco analítico que facilita la comprensión de su aplicabilidad en diferentes contextos de servicios cloud. El enfoque descriptivo adoptado resulta particularmente valioso para proporcionar una base sólida que permita a investigadores y profesionales seleccionar las técnicas más apropiadas según sus necesidades específicas, contribuyendo así al avance tanto teórico como práctico en el campo de la detección de anomalías en entornos cloud.

DESARRROLLO

El aprendizaje no supervisado representa una aproximación metodológica fundamental en el campo de la inteligencia artificial y el análisis de datos, caracterizada por su capacidad para identificar estructuras inherentes sin requerir etiquetado previo de información, esta técnica se distingue por su potencial para descubrir patrones ocultos, relaciones complejas y agrupaciones naturales dentro de conjuntos de datos multidimensionales (Almuqati et al., 2024), en el contexto de servicios cloud computing, el aprendizaje no supervisado se configura como una herramienta estratégica para comprender comportamientos dinámicos y eventos atípicos en infraestructuras tecnológicas altamente complejas, la esencia de esta metodología radica en su capacidad para generar insights sin intervención humana directa, utilizando algoritmos que exploran la estructura intrínseca de los datos mediante técnicas de clustering, reducción dimensional y detección de anomalías.

Consecuentemente, los servicios cloud computing emergen como ecosistemas tecnológicos distribuidos que demandan aproximaciones innovadoras para gestionar su complejidad inherente, la infraestructura cloud se caracteriza por su naturaleza escalable, elástica y altamente dinámica; lo que genera desafíos significativos en términos de monitoreo, seguridad y optimización de recursos computacionales, en este contexto, la detección de anomalías se configura como un proceso crítico para identificar comportamientos desviados que puedan comprometer la integridad, rendimiento y seguridad de los sistemas (Almuqati et al., 2024), las técnicas de aprendizaje no supervisado proporcionan un marco metodológico avanzado para analizar patrones complejos, identificar tendencias emergentes y desarrollar estrategias proactivas de gestión de infraestructura tecnológica.

Las anomalías en los servicios cloud computing representan desviaciones significativas de las normas establecidas, lo que plantea riesgos para la seguridad y la eficiencia operativa (Liu et al., 2024), desde una perspectiva técnica, las anomalías pueden manifestarse en múltiples dimensiones, incluyendo variaciones en consumo de recursos, patrones de tráfico atípicos, desviaciones en métricas de rendimiento y comportamientos estadísticamente significativos que no siguen los modelos predictivos tradicionales, la identificación precisa de estas anomalías requiere el desarrollo de algoritmos sofisticados capaces de distinguir entre variaciones naturales y eventos realmente críticos, considerando la complejidad y heterogeneidad de los entornos cloud computing contemporáneos (Demirbaga, 2024).

Por consiguiente, los algoritmos de clustering se presentan como herramientas fundamentales en el proceso de detección de anomalías, permitiendo agrupar elementos con características similares y resaltar aquellos que no se ajustan a los patrones predominantes (Oyelade et al., 2019), técnicas como K-means, DBSCAN, Gaussian Mixture Models y algoritmos jerárquicos proporcionan aproximaciones matemáticas para segmentar datos multidimensionales, identificando estructuras subyacentes y elementos estadísticamente divergentes, la selección del algoritmo de clustering más apropiado depende de múltiples factores, incluyendo la naturaleza del conjunto de datos, dimensionalidad, distribución estadística y objetivos específicos de análisis en el contexto de servicios cloud computing (Velunachiyar & Sivakumar, 2023).

Una síntesis conceptual que facilite la comprensión de términos especializados en detección de anomalías se presenta a continuación en la Tabla 1. Glosario de Términos Clave en Detección de Anomalías.

Tabla 1. Glosario de Términos Clave en Detección de Anomalías.

Término	Definición	Relevancia
Aprendizaje No Supervisado	Técnica de machine learning que identifica patrones sin etiquetado previo de datos, explorando estructuras intrínsecas mediante agrupamiento y reducción dimensional (Almuqati et al., 2024).	Metodología fundamental para descubrir patrones de anomalías sin necesidad de etiquetas previas en entornos cloud.
Anomalía en Servicios Cloud	Desviación estadísticamente significativa del comportamiento normal en infraestructuras computacionales distribuidas (Liu et al., 2024).	Objeto central de estudio, representa el fenómeno que se busca identificar y caracterizar
Clustering	Técnica de agrupamiento que segmenta datos en conjuntos con características similares (Oyelade et al., 2019).	Método clave para agrupar y distinguir comportamientos típicos de anomalías en servicios cloud
Servicios Cloud Computing	Infraestructura tecnológica distribuida que proporciona recursos computacionales bajo demanda (Huawei Technologies Co., 2025).	Contexto tecnológico específico de aplicación de las técnicas de detección de anomalías
Detección de Anomalías	Proceso de identificación de patrones que divergen significativamente de un comportamiento esperado (Almuqati et al., 2024).	Objetivo principal de la investigación para identificar desviaciones en infraestructuras cloud.
Algoritmos No Supervisados	Métodos computacionales que extraen información sin requerir etiquetado previo de datos (Almuqati et al., 2024).	Técnicas de análisis para explorar patrones de anomalías sin intervención manual.
Entropía Computacional	Medida de incertidumbre y complejidad en sistemas de procesamiento de información (Vadhan, 2019).	Indicador para cuantificar la variabilidad y comportamientos atípicos en servicios cloud.
Reducción Dimensional	Técnica que simplifica datos manteniendo características fundamentales (Almuqati et al., 2024).	Método para simplificar y caracterizar conjuntos de datos complejos de servicios cloud.
Infraestructura Tecnológica	Conjunto de recursos computacionales interconectados que soportan servicios digitales (Kraus et al., 2022).	Ambiente de análisis y aplicación de técnicas de detección de anomalías.
Patrón Atípico	Comportamiento estadísticamente divergente de la norma establecida (Almuqati et al., 2024).	Elemento focal para identificar y caracterizar anomalías en entornos cloud.

La Teoría de Sistemas Complejos constituye un marco fundamental para comprender la dinámica de servicios cloud computing, proporcionando una perspectiva holística sobre la interacción de componentes tecnológicos y su comportamiento emergente (Karaca, 2022), esta teoría postula que los sistemas complejos exhiben propiedades no lineales, auto organización y capacidad de adaptación, elementos críticos para interpretar la evolución de infraestructuras tecnológicas distribuidas, la aplicación de principios de sistemas complejos en el contexto de aprendizaje no supervisado permite desarrollar modelos predictivos más robustos, capaces de capturar la naturaleza dinámica y evolutiva de los entornos cloud computing .

Seguidamente, la **Teoría de la Información** inicialmente desarrollada por Claude Shannon, establece las bases matemáticas para cuantificar la incertidumbre y complejidad en sistemas de comunicación y procesamiento de datos, en el contexto de detección de anomalías, esta teoría proporciona herramientas para evaluar la entropía de los datos, identificando desviaciones significativas que pueden representar eventos atípicos o potenciales amenazas en infraestructuras tecnológicas, la medición de la información mutua (Il-agure & Attallah, 2019) y la divergencia de Kullback-Leibler (Tekeoglu et al., 2019) se configuran como técnicas avanzadas para analizar la distribución estadística de datos en servicios cloud, facilitando la identificación de patrones anómalos con mayor precisión.

La **Teoría de redes complejas** proporciona un marco fundamental para el análisis de infraestructuras de computación en la nube, modelando las complejas interrelaciones entre los componentes tecnológicos, este enfoque permite identificar nodos críticos, patrones de conectividad y vulnerabilidades dentro del sistema, los algoritmos derivados de la teoría de redes son esenciales para la detección de anomalías, ya que facilitan el análisis de grafos mediante el uso de métricas como la centralidad, la modularidad y el coeficiente de agrupamiento, estas métricas son cruciales para desentrañar la estructura subyacente de los servicios en la nube, permitiendo una comprensión más profunda de su funcionamiento y mejorando la resiliencia del sistema frente a posibles fallos o ataques (Liu et al., 2022).

La **Teoría de Aprendizaje Estadístico** proporciona fundamentos matemáticos para desarrollar modelos predictivos que generalicen efectivamente a partir de conjuntos

de datos limitados (Huang et al., 2018), esta teoría aborda los desafíos de sobreajuste y subajuste en algoritmos de machine learning, estableciendo principios para seleccionar modelos que capturen la complejidad intrínseca de los datos sin perder capacidad de generalización, en el contexto de detección de anomalías, los principios de regularización, validación cruzada y selección de modelos se configuran como herramientas fundamentales para garantizar la robustez de los algoritmos de aprendizaje no supervisado, estos principios permiten desarrollar modelos que no solo se ajustan a los datos de entrenamiento, sino que también demuestran una capacidad predictiva consistente en conjuntos de datos no vistos (Zhouet al., 2024).

La **Teoría de Conjuntos Difusos** representa una aproximación matemática para manejar la incertidumbre y ambigüedad inherente en sistemas complejos, superando las limitaciones de los enfoques binarios tradicionales (Sissodia et al., 2025), esta teoría permite modelar conceptos y relaciones que no pueden ser definidos con precisión absoluta, resultando especialmente útil en la detección de anomalías donde los límites entre comportamientos normales y atípicos no son completamente discretos, por otro lado, los algoritmos de lógica difusa utilizan variables lingüísticas y funciones de pertenencia para representar valores, lo que permite la interpretación de datos complejos y captura la naturaleza gradual de las anomalías en los servicios de computación en nube, mejorando así los procesos de toma de decisiones en la administración de recursos.

Consecuentemente, la **Teoría de la Complejidad Computacional** analiza el poder intrínseco y las limitaciones de los recursos computacionales como el tiempo, el espacio y la aleatoriedad, estableciendo límites fundamentales en el procesamiento de la información (Tian et al., 2021), en el contexto de detección de anomalías, esta teoría permite evaluar la eficiencia computacional de diferentes algoritmos de aprendizaje no supervisado, considerando métricas como tiempo de procesamiento, consumo de memoria y escalabilidad, la comprensión de la complejidad computacional resulta crítica para seleccionar estrategias de detección de anomalías que balanceen precisión y rendimiento en entornos cloud dinámicos.

La **Teoría de Probabilidad y Procesos Estocásticos** proporciona herramientas matemáticas esenciales para modelar sistemas caracterizados por la aleatoriedad y la incertidumbre, siendo fundamental en la detección de anomalías en los servicios de computación en la nube, esta teoría permite desarrollar modelos probabilísticos que capturan la variabilidad inherente de los sistemas tecnológicos, distinguiendo entre variaciones naturales y eventos verdaderamente atípicos, lo que es crucial para mantener la integridad y seguridad de los sistemas en entornos cloud (Qi et al., 2022), las distribuciones de probabilidad, procesos de Markov y técnicas de inferencia estadística se configuran como aproximaciones avanzadas para identificar comportamientos desviados con rigor matemático.

Adicionalmente, la **Teoría de sistemas adaptativos y de control** proporciona un marco sólido para comprender y administrar sistemas complejos que evolucionan dinámicamente, particularmente en el contexto de infraestructuras de nube resilientes, esta teoría hace hincapié en la autoorganización, la adaptación y el aprendizaje, que son esenciales para que los sistemas respondan eficazmente a los cambios ambientales (Pham & Kaneko, 2024), los mecanismos de retroalimentación, autorregulación y aprendizaje continuo se configuran como elementos centrales para desarrollar estrategias de detección de anomalías que no solo identifiquen eventos atípicos, sino que también faciliten la adaptación proactiva de los servicios cloud.

Análisis Metodológico de Algoritmos de Aprendizaje No Supervisado para Detección de Anomalías en Servicios Cloud

La detección de anomalías en servicios cloud es una tarea crítica para garantizar la seguridad y el rendimiento de las infraestructuras tecnológicas, los algoritmos de aprendizaje no supervisado han demostrado ser herramientas eficaces para identificar comportamientos atípicos sin la necesidad de datos etiquetados, uno de los algoritmos más destacados en la detección de anomalías es el **Isolation Forest (IF)**, que se basa en la construcción de árboles de decisión para aislar puntos de datos, este método ha sido ampliamente estudiado y aplicado debido a su capacidad para manejar grandes volúmenes de datos y su eficiencia computacional (Bouman et al., 2023), estudios recientes han demostrado que el IF es particularmente eficaz en la detección de anomalías en datos de alta dimensionalidad, lo que lo hace ideal para entornos cloud.

Otro algoritmo relevante es el **k-Nearest Neighbors (kNN)**, que identifica anomalías basándose en la densidad local de los datos, este enfoque es útil para detectar anomalías locales que pueden no ser evidentes en un análisis global, la simplicidad y efectividad del kNN lo han convertido en una opción popular en la investigación y aplicación práctica de la detección de anomalías en servicios cloud (Bouman et al., 2023), sin embargo, uno de los desafíos de kNN es su sensibilidad a la elección del parámetro k y la necesidad de un cálculo intensivo de distancias, lo que puede ser computacionalmente costoso en grandes conjuntos de datos, a pesar de esto, su capacidad para adaptarse a diferentes tipos de datos y su robustez en la detección de anomalías lo hacen una herramienta

valiosa. Estudios recientes han investigado variantes del kNN, como el uso de técnicas de reducción de dimensionalidad para mejorar su eficiencia y precisión.

El **Autoencoder**, una red neuronal diseñada para aprender una representación comprimida de los datos, también ha sido ampliamente utilizado en la detección de anomalías, son capaces de capturar patrones complejos en los datos y detectar desviaciones significativas de estos patrones, su capacidad para manejar datos no lineales y su flexibilidad los hacen adecuados para una variedad de aplicaciones en servicios cloud (Kadiyala et al., 2022), además, pueden ser entrenados en datos no etiquetados, lo que los hace ideales para escenarios donde la disponibilidad de datos etiquetados es limitada, la capacidad de los autoencoders para aprender representaciones latentes de los datos también permite una detección más precisa y robusta de las anomalías.

La **Análisis de Componentes Principales (PCA)** es otra técnica comúnmente utilizada en la detección de anomalías, reduce la dimensionalidad de los datos, facilitando la identificación de puntos de datos que se desvían significativamente de la estructura principal, aunque PCA es una técnica lineal, su simplicidad y efectividad la mantienen relevante en el campo de la detección de anomalías (Kadiyala et al., 2022), sin embargo, uno de los desafíos de PCA es su suposición de linealidad, que puede no ser adecuada para todos los tipos de datos, a pesar de esto, PCA sigue siendo una herramienta valiosa debido a su capacidad para simplificar datos complejos y mejorar la interpretabilidad de los resultados.

El **Clustering basado en densidad**, como el algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise), es eficaz para identificar anomalías en conjuntos de datos con estructuras complejas, DBSCAN puede detectar grupos de datos de densidad variable y aislar puntos de datos que no pertenecen a ningún grupo, identificándolos como anomalías, este enfoque es particularmente útil en la detección de anomalías en datos espaciales y temporales, donde las relaciones entre los datos son cruciales, además, DBSCAN no requiere especificar el número de clusters de antemano, lo que lo hace más flexible y adaptable a diferentes tipos de datos y escenarios (Bouman et al., 2023), investigaciones recientes han explorado variantes de DBSCAN, como HDBSCAN, que mejoran su capacidad para manejar datos con estructuras aún más complejas.

La **Máquina de Soporte Vectorial (SVM)** en su variante de una clase (One-Class SVM) es otra técnica utilizada para la detección de anomalías, este algoritmo aprende una frontera que separa los datos normales de los anómalos, siendo particularmente útil en escenarios donde las anomalías son raras y difíciles de etiquetar, la SVM de una clase ha demostrado ser efectiva en la detección de anomalías en datos de red y sistemas cloud, aunque, uno de los desafíos de SVM es su sensibilidad a la elección de parámetros y la necesidad de un ajuste cuidadoso para obtener resultados óptimos, a pesar de esto, su capacidad para manejar datos de alta dimensionalidad y su robustez en la detección de anomalías la hacen una herramienta valiosa (Bouman et al., 2023).

El **Modelo de Mezcla Gaussiana (GMM)** es un enfoque probabilístico que modela la distribución de los datos como una combinación de múltiples distribuciones gaussianas, es útil para detectar anomalías al identificar puntos de datos que tienen una baja probabilidad de pertenecer a cualquiera de las distribuciones modeladas, este enfoque es especialmente útil en la detección de anomalías en datos continuos y de alta dimensionalidad, sin embargo, uno de los desafíos de GMM es su suposición de que los datos siguen una distribución gaussiana, lo que puede no ser adecuado para todos los tipos de datos (Lu et al., 2024).

El **Algoritmo de Clustering K-means** también se utiliza en la detección de anomalías, aunque su aplicación es más limitada en comparación con otros métodos. K-means agrupa los datos en k clusters y los puntos de datos que están lejos de cualquier centro de cluster se consideran anomalías. Este método es simple y eficiente, pero puede no ser adecuado para datos con estructuras complejas (Ahmed-Rana, 2024), además, K-means requiere especificar el número de clusters de antemano, lo que puede ser un desafío en escenarios donde el número óptimo de clusters no es conocido, a pesar de esto, K-means sigue siendo una herramienta popular debido a su simplicidad y eficiencia.

La **Red Neuronal de Crecimiento Competitivo (GNG)** es una técnica menos común pero efectiva para la detección de anomalías, es una red neuronal que se adapta dinámicamente a la estructura de los datos, permitiendo la identificación de anomalías a través de la evolución de la red, este enfoque es particularmente útil en la detección de anomalías en datos no estacionarios, sin embargo, uno de los desafíos de GNG es su complejidad y la necesidad de un ajuste cuidadoso de los parámetros para obtener resultados óptimos, aun así, su capacidad para adaptarse a diferentes tipos de datos y su robustez en la detección de anomalías la hacen una herramienta valiosa (Alalkawı et al., 2023).

El **Algoritmo Subespacial de Anomalías (SSA)** es una técnica avanzada que combina la reducción de dimensionalidad con la detección de anomalías, identifica subespacios en los que las anomalías son más evidentes, mejorando la precisión de la detección, este enfoque es útil en la detección de anomalías en datos de alta dimensionalidad y complejidad, sin embargo, uno de los desafíos de SSA es su complejidad y la necesidad de un ajuste cuidadoso de los parámetros para obtener resultados óptimos, a pesar de esto, su capacidad para mejorar

la precisión de la detección de anomalías lo hace una herramienta valiosa en la investigación y aplicación práctica (Mandrikova et al., 2023).

Finalmente, el **Algoritmo de Detección de Anomalías Basado en Gráficos (GAD)** utiliza estructuras de grafos para modelar las relaciones entre los datos y detectar anomalías, es eficaz en la detección de anomalías en datos de red y sistemas distribuidos, donde las relaciones entre los datos son cruciales, lo que ha demostrado ser particularmente útil en la detección de anomalías en servicios cloud, donde la interconexión y la dependencia entre los componentes del sistema pueden presentar desafíos significativos, pero uno de los desafíos de GAD es su complejidad y la necesidad de un ajuste cuidadoso de los parámetros para obtener resultados óptimos, a pesar de esto, su capacidad para modelar relaciones complejas y mejorar la precisión de la detección de anomalías lo hace una herramienta valiosa (Ekle & Eberle, 2024).

La detección de anomalías en servicios cloud es esencial para garantizar la seguridad y el rendimiento en entornos tecnológicos complejos la Tabla 2 sintetiza las principales características, ventajas y desafíos de los algoritmos antes detallados:

Tabla 2. Análisis Metodológico de Algoritmos de Aprendizaje No Supervisado para Detección de Anomalías en Servicios Cloud.

Algoritmo	Ventajas	Desafíos
Isolation Forest (IF)	- Eficiente computacionalmente - Ideal para datos de alta dimensionalidad	- Menor precisión en datos con relaciones complejas
k-Nearest Neighbors (kNN)	- Detecta anomalías locales - Adaptable a distintos tipos de datos	- Sensible al parámetro k - Alto costo computacional
Autoencoder	- Captura patrones complejos - Ideal para datos no lineales - Requiere pocos datos etiquetados	- Requiere gran capacidad de cómputo - Mayor complejidad en implementación
PCA	- Simplifica datos complejos - Fácil interpretación	- Suposición de linealidad - Menor precisión en datos no lineales
DBSCAN	- Detecta estructuras complejas - No requiere predefinir el número de clusters	- Sensible a parámetros de densidad - No ideal para datos con ruido extremo
One-Class SVM	- Eficaz en datos de alta dimensionalidad - Útil para escenarios con pocas anomalías	- Sensible al ajuste de parámetros - Alto costo computacional
Modelo de Mezcla Gaussiana (GMM)	-Útil en datos continuos - Modela distribuciones probabilísticas	- Asume distribución gaussiana - Menor eficacia en datos con estructuras complejas
K-means	- Simplicidad y eficiencia	- Necesita predefinir el número de clusters - Menor precisión en datos con alta complejidad
Red Neuronal de Crecimiento Competitivo (GNG)	- Adaptable a datos no estacionarios - Robusta en detección de anomalías	- Ajuste complejo de parámetros - Alta complejidad computacional
SSA	- Precisión mejorada para alta dimensionalidad	- Ajuste complejo - Alta demanda de recursos
GAD	- Modela relaciones complejas - Útil en sistemas distribuidos	- Alta complejidad computacional - Ajuste cuidadoso de parámetros

Para complementar el marco teórico y proporcionar evidencia empírica sobre la aplicación práctica de las Técnicas de Aprendizaje No Supervisado para Detección de Anomalías en Servicios Cloud, es pertinente analizar casos de estudio que demuestren la implementación exitosa.

Caso #1:

Arquitectura en Tiempo Real para la Detección de Fraude en Transacciones Digitales (Hanae et al., 2023)

- **Objetivo:** Proponer una arquitectura en tiempo real para detectar fraudes en transacciones digitales utilizando análisis de comportamiento y técnicas de big data.
- **Métodos:** Combinación de herramientas como Spark, Kafka y el algoritmo de aprendizaje automático Isolation Forest (IF) para la detección de anomalías.
- **Resultados:** Precisión del 99% y precisión del 87% en la detección de fraudes en un conjunto de datos significativo.

Caso #2:

Detección de anomalías utilizando K-Means y LSTM para el mantenimiento predictivo de plantas fotovoltaicas a gran escala (Zulfauzi et al., 2023)**.**

- **Objetivo:** Mejorar la detección de fallos en plantas solares mediante técnicas de aprendizaje automático.
- **Métodos:** Se utilizó K-Means para agrupar datos y LSTM para detectar anomalías en la corriente eléctrica.
- **Resultados:** LSTM mostró mayor precisión y menor error relativo en comparación con redes neuronales artificiales (ANN).

- **Conclusión:** La combinación de K-Means y LSTM optimiza el mantenimiento predictivo, reduciendo costos y mejorando la eficiencia.

Caso #3:

SAGE – Habilitando la Depuración Práctica del Rendimiento en la Nube con Aprendizaje No Supervisado (Gan et al., 2022)

- **Objetivo:** Desarrollar una solución eficaz para identificar y solucionar problemas de rendimiento en sistemas complejos de microservicios en la nube, el enfoque es utilizar técnicas de aprendizaje automático no supervisado para analizar grandes cantidades de datos de rendimiento sin la necesidad de una etiquetación manual previa.
- **Métodos:** Este sistema utiliza dos técnicas clave de aprendizaje no supervisado:
  - **Redes Bayesianas Causales:** Estas redes permiten modelar las relaciones de causa y efecto entre diferentes componentes del sistema, lo que ayuda a identificar qué microservicio está causando un problema específico.
  - **Autoencoders Variacionales Gráficos:** Estas redes reconstruyen los datos de entrada y detectan anomalías que pueden indicar problemas de rendimiento.
- **Resultados:** Los resultados del estudio demuestran que Sage es altamente efectivo en la identificación de las causas raíz de los problemas de rendimiento en sistemas de microservicios, al analizar los datos de rendimiento de manera no supervisada, SAGE puede:
  - **Identificar con precisión los microservicios problemáticos:** Al modelar las relaciones causales entre los componentes del sistema, Sage puede señalar directamente al culpable de un problema de rendimiento.
  - **Detectar anomalías en los datos:** Los autoencoders variacionales gráficos permiten a Sage identificar patrones inusuales en los datos que pueden indicar problemas subyacentes.
  - **Reducir el tiempo de resolución de problemas:** Al automatizar el proceso de identificación de problemas, Sage permite a los ingenieros de software resolver los problemas de rendimiento de manera más rápida y eficiente.
  - **Conclusión:** El estudio concluye que el aprendizaje no supervisado es una herramienta poderosa para la depuración del rendimiento en sistemas de microservicios en la nube, SAGE, demuestra que es posible automatizar la identificación de problemas de rendimiento sin la necesidad de una intervención humana significativa, al utilizar técnicas de aprendizaje automático avanzadas, puede ayudar a las organizaciones a mejorar la confiabilidad y el rendimiento de sus aplicaciones en la nube.

El análisis metodológico exhaustivo de las técnicas de aprendizaje no supervisado para la detección de anomalías en servicios cloud ha revelado patrones significativos en términos de efectividad y aplicabilidad, la investigación de las fuentes bibliográficas demuestra una clara evolución en la sofisticación de los algoritmos, desde métodos tradicionales como K-means hasta aproximaciones más avanzadas como Autoencoders y SAGE, los resultados del primer caso de estudio documentan una precisión del 99% y una exactitud del 87% en la detección de fraudes mediante la implementación de Isolation Forest en conjunto con herramientas de procesamiento distribuido como Spark y Kafka, evidenciando la efectividad de los enfoques híbridos en entornos cloud.

El análisis de los casos de estudio revela una tendencia hacia la implementación de soluciones que combinan múltiples técnicas de aprendizaje no supervisado. El caso del sistema SAGE demuestra la efectividad de combinar Redes Bayesianas Causales con Autoencoders Variacionales para la identificación de problemas de rendimiento en sistemas de microservicios. El segundo caso de estudio confirma que la combinación de K-Means y LSTM resulta más efectiva que las redes neuronales artificiales tradicionales para la detección de anomalías en el mantenimiento predictivo de plantas fotovoltaicas.

Los resultados indican que la efectividad de los algoritmos varía según el contexto específico de aplicación, se evidencia también, que técnicas como DBSCAN son particularmente efectivas para datos con estructuras complejas, mientras que Isolation Forest destaca en el manejo de datos de alta dimensionalidad, esta variabilidad en el rendimiento subraya la importancia de seleccionar los algoritmos en función de las características específicas del entorno cloud y los objetivos particulares de la detección de anomalías.

CONCLUSIONES

La detección de anomalías en servicios cloud utilizando técnicas de aprendizaje no supervisado representa un campo de investigación dinámico y en constante evolución, los algoritmos discutidos en este análisis metodológico proporcionan una base sólida para el desarrollo de sistemas de detección de anomalías robustos y eficientes, la selección del algoritmo adecuado depende de las características específicas de los datos y del entorno de aplicación, lo que subraya la importancia de un análisis bibliográfico comprehensivo y actualizado, esta perspectiva fundamental se alinea con los hallazgos del análisis metodológico realizado, donde algoritmos como Isolation Forest, Autoencoders y DBSCAN han demostrado capacidades excepcionales para identificar patrones anómalos sin requerir etiquetado previo de datos, característica esencial en entornos cloud dinámicos y complejos.

La efectividad de estos algoritmos varía según el contexto de aplicación, donde factores como la dimensionalidad de los datos, la escalabilidad del sistema y los requisitos de rendimiento juegan un papel crucial en la selección de la técnica más apropiada, respondiendo así al primer objetivo específico de la investigación, el análisis exhaustivo de casos de estudio ha proporcionado evidencia empírica sólida sobre la aplicación exitosa de técnicas de aprendizaje no supervisado en entornos cloud, cumpliendo con el segundo objetivo específico de la investigación.

Los casos analizados, incluyendo la arquitectura en tiempo real para la detección de fraude en transacciones digitales y el sistema SAGE para la depuración del rendimiento en la nube, demuestran que la implementación de estas técnicas puede alcanzar niveles de precisión notables en la detección de anomalías, especialmente cuando se emplean enfoques híbridos que combinan múltiples algoritmos, estos resultados validan la aplicabilidad práctica de las metodologías estudiadas y su capacidad para mejorar significativamente la seguridad y eficiencia de los servicios cloud, proporcionando evidencia tangible de su efectividad en escenarios del mundo real.

De manera general, se concluye que la detección efectiva de anomalías en servicios cloud requiere un enfoque adaptativo que combine múltiples técnicas de aprendizaje no supervisado, la investigación demuestra que no existe una solución única que se adapte a todos los escenarios, sino que la efectividad depende de una selección cuidadosa de algoritmos basada en las características específicas del entorno cloud y los requisitos particulares de cada implementación, esta conclusión tiene implicaciones significativas para el desarrollo futuro de sistemas de detección de anomalías, sugiriendo la necesidad de marcos de trabajo flexibles que puedan adaptarse a la naturaleza dinámica y evolutiva de las infraestructuras cloud computing.

Las limitaciones identificadas en los métodos actuales y los desafíos emergentes en el campo de la detección de anomalías señalan direcciones prometedoras para investigaciones futuras, se recomienda profundizar en el desarrollo de técnicas híbridas que combinen el aprendizaje no supervisado con otros enfoques de inteligencia artificial, así como en la optimización de algoritmos para mejorar su eficiencia computacional en entornos cloud de gran escala, adicionalmente, resulta fundamental explorar la integración de técnicas de procesamiento en tiempo real y la adaptación dinámica de parámetros para mejorar la capacidad de respuesta ante anomalías emergentes, estas recomendaciones establecen una base sólida para futuras investigaciones en el campo, contribuyendo al desarrollo continuo de soluciones más efectivas para la detección de anomalías en servicios cloud.

REFERENCIAS BIBLIOGRÁFICAS

Acosta-Servín, S., Veytia-Bucheli, M. G., & Cáceres-Mesa, M. L. (2025). Innovar en la práctica docente. Desarrollo de competencias digitales en la Licenciatura. Sophia Editions.

Ahmed-Rana, D. U. (2024). Application of Data Mining Combined with K-means Clustering Algorithm in Enterprises' Risk Audit. ***Qeios***. https://doi.org/10.32388/G9G0S3

Alalkawı, M. D., Shehabı, S. A., & Imamoglu, M. Y. (2023). PTGNG: An Evolutionary Approach for Parameter Optimization in the Growing Neural Gas Algorithm. ***International Journal of Computational and Experimental Science and Engineering, 9***(2), 91-101. https://doi.org/10.22399/ijcesen.1282146

Almuqati, M., Sidi, F., Mohd, S., Zolkepli, M., & Ishak, I. (2024). Challenges in Supervised and Unsupervised Learning: A Comprehensive Overview. ***International Journal on Advanced Science Engineering and Information, 14***(4). https://doi.org/10.18517/ijaseit.14.4.20191

Bouman, R., Bukhsh, Z., & Heskes, T. (2023). Unsupervised anomaly detection algorithms on real-world data: how many do we need? ***Journal of Machine Learning Research***, 25. 1-34. https://doi.org/10.48550/arXiv.2305.00735

Casimiro-Urcos, W. H., Casimiro-Urcos, C. N., Quinteros-Osorio, R. O., Tello-Conde, A. R., & Casimiro-Guerra, G. (2025). Docentes conectados: Evaluando las competencias digitales en la Educación Superior. Sophia Editions.

Demirbaga, U. (2024). Advancing anomaly detection in cloud environments with cutting-edge generative AI for expert systems. ***Wiley***.

Ekle, O. A., & Eberle, W. (2024). Anomaly Detection in Dynamic Graphs: A Comprehensive Survey. ***ACM Transactions on Knowledge Discovery from Data, 18***(8). https://doi.org/10.1145/3669906

Gan, Y., Liang, M., Dev, S., Lo, D., & Delimitrou, C. (2022). Enabling Practical Cloud Performance Debugging with Unsupervised Learning. ***ACM SIGOPS Operating Systems Review, 56***(1), 34-41. https://doi.org/10.1145/3544497.3544503

Hanae, A., Abdellah, B., Saida, E., & Youssef, G. (2023). End-to-End Real-time Architecture for Fraud Detection in Online Digital Transactions. ***International Journal of Advanced Computer Science and Applications, 14***(6). https://doi.org/10.14569/ijacsa.2023.0140680

Huang, W., Li, G.-M., & Chen, W.-W. (2018). A Review of Statistical Learning Theory. ***DEStech Transactions on Engineering and Technology Research***. https://www.doi.org/10.12783/DTETR/PMSMS2018/24953

Huawei Technologies Co., L. (2025). ***Cloud Computing Technology.*** Springer. https://doi.org/10.1007/978-981-19-3026-3

Il-agure, Z., & Attallah, B. (2019). How mutual information interprets anomalies using different clustering. ***International Journal of Grid and Utility Computing***. https://www.doi.org/10.1504/IJGUC.2019.10018229

Jurado Pruna, F., Valeria Yarad, J. P., & Carrión Jumbo, J. L. (2020). Análisis de las características del sector microempresarial en latinoamérica y sus limitantes en la adopción de tecnologías para la seguridad de la información. ***Revista Científica Ecociencia, 7***(1), 1-26. https://doi.org/10.21855/ecociencia.71.303

Kadiyala, P., Shanmukhasai, K. V., Budem, S. S., & Maddikunta, P. K. (2022). Anomaly Detection Using Unsupervised Machine Learning Algorithms. En, A. Makkar y N. Kumar, Deep Learning for Security and Privacy Preservation in IoT. (pp. 113–125). Springer.

Karaca, Y. (2022). Chapter 2 - Theory of complexity, origin and complex systems. En, Y. Karaca, B. Dumitru, Z. Yu-Dong, O. Gervasi, & M. Moonis, ***Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems.*** (pp. 9-20). Academic Press.

Kraus, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. (2022). Digital transformation in business and management research: An overview of the current status quo. ***International Journal of Information Management, 63***. https://doi.org/10.1016/j.ijinfomgt.2021.102466

Lavado-Rojas, B. M., Pomahuacre-Gómez, W., Castro-Fernández, M. A., Castellano-Inga, A. F., Zárate-Aliaga, E. C., & López-Torres, M. (2025). ***Competencias digitales y lenguas extranjeras: Un enfoque formativo para la educación universitaria***. Sophia Editions.

Li, H. (2023). ***Machine Learning Methods***. Springer.

Liu, S., Zhou, Y., Ying, L., Tian, Y., Zhang, J., Zhou, S., Cui, W., Lin, Q., Moscibroda, T., Zhang, H., Weng, D., & Wu, Y. (2024). RCInvestigator: Towards Better Investigation of Anomaly Root Causes in Cloud Computing Systems. https://doi.org/10.48550/arXiv.2405.15571

Liu, Y., Zhang, Y., & & Wang, X. (2022). A survey on anomaly detection in cloud computing: Techniques and applications. ***Journal of Cloud Computing: Advances, Systems and Applications, 11***(1), 1-21. https://doi.org/10.1186/s13677-022-00261-4

Lu, W., Ding, D., Wu, F., & Yuan, G. (2024). An Efficient Gaussian Mixture Model and Its Application to Neural Network. ***Preprints***. https://doi.org/10.20944/preprints202302.0275.v3

Mandrikova, O., Mandrikova, B., & Esikov, O. (2023). Detection of Anomalies in Natural Complicated Data Structures Based on a Hybrid Approach. ***Mathematics, 11***(11). https://doi.org/https://doi.org/10.3390/math11112464

Oyelade, J., Isewon, I., Oladipupo, O., Emebo, O., Omogbadegun, Z., & Aromolaran, O. (2019). Data Clustering: Algorithms and Its Applications. (Ponencia). ***19th International Conference on Computational Science and Its Applications (ICCSA)***, St. Petersburg, Russia.

Pham, T. M., & Kaneko, K. (2024). Dynamical theory for adaptive systems. ***Journal of Statistical Mechanics: Theory and Experiment***. https://doi.org/10.1088/1742-5468/ad8223

Priarone, A., Albertin, U., Cena, C., Martini, M., & Chiaberge, M. (2024). Unsupervised Novelty Detection Methods Benchmarking with Wavelet Decomposition. https://doi.org/10.48550/arXiv.2409.07135

Qi, B., Zhang, P., Wu, H., & Yan, M. (2022). Cloud Resource Scheduling Method based on Markov Process and the Cuckoo Search. ***Journal of Physics: Conference Series***. https://doi.org/10.1088/1742-6596/2320/1/012030

Saikiran, N., Reddy, K. Y., Reddy, C. P., & Karthik, S. (2024). Advanced Anomaly Detection in Cloud Security Using Gini Impurity and ML. (Ponencia). ***3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)***. Salem, India.

Sissodia, R., Rauthan, M. S., Barthwal, V., & Dwivedi, V. (2025). Fuzzy Logic. En, S. Aouadni y I. Aouadni, ***Recent Theories and Applications for Multi-Criteria Decision-Making.*** (pp. 279-310). IGI Global Scientific Publishing.

Tekeoglu, A., Andriamanalimanana, B., Bekiroglu, K., Sengupta, S., Chiang, C.-F., & Reale, M. (2019). Symmetric kullback-leibler divergence of softmaxed distributions for anomaly scores. (Ponencia). ***IEEE Conference on Communications and Network Security (CNS)***. Washington, DC, USA

Tian, C., Plank, J. S., Hurst, B., & Zhou, R. (2021). Computational Techniques for Investigating Information Theoretic Limits of Information Systems. ***Information an International Interdisciplinary Journal, 12***(2). https://doi.org/10.3390/info12020082

Vadhan, S. (2019). Computational entropy. En, O. Goldreich, ***Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali.*** (pp. 693 - 726). ACM Books.

Velunachiyar, S., & Sivakumar, K. (2023). Some Clustering Methods, Algorithms and their Applications. ***International Journal on Recent and Innovation Trends in Computing and Communication, 11***(6S), 401–410. https://doi.org/10.17762/ijritcc.v11i6s.6946

Zhou, T.-Y., Lau, M., Chen, J., Lee, W., & Huo, X. (2024). Optimal Classification-based Anomaly Detection with Neural Networks: Theory and Practice. ***arXiv***. https://doi.org/10.48550/arXiv.2409.08521

Zulfauzi, I. A., Dahlan, N. Y., Sintuya, H., & Setthapun, W. (2023). Anomaly detection using K-Means and long-short term memory for predictive maintenance of large-scale solar (LSS) photovoltaic plant. ***Energy Reports, 9***(12), 154-158. https://doi.org/https://doi.org/10.1016/j.egyr.2023.09.159