GADE. REV. CIENT. VOL. 6. NÚM. 1 (2026)
ISSN: 2745-2891

Estudio comparativo de técnicas en la inteligencia artificial explicable basados en métodos de explicabilidad para datos estructurados

Comparative study of techniques in explainable artificial intelligence based on explainability methods for structured data

Fausto Raúl Orozco Lara*
fausto.orozcol@ug.edu.ec
Mariela Paola Espinoza Martínez*
mariela.espinozam@ug.edu.ec
Mell Naomi Mainato Acosta*
mell.mainatoa@ug.edu.ec
Diego Alexander Siguencia Ordóñez*
diego.siguenciao@ug.edu.ec
*Universidad de Guayaquil, Ecuador.

Recibido: 17/12/2025-Aceptado:24/02/2026
Correspondencia: fausto.orozcol@ug.edu.ec

Resumen

Este artículo, derivado de una investigación sobre la Inteligencia Artificial Explicable (XAI) para la detección de actividades inusuales en aulas virtuales, aborda los desafíos críticos de seguridad y honestidad académica surgidos tras la migración universitaria a entornos digitales. Ante la inviabilidad de la supervisión manual sobre el volumen masivo de datos generado, el estudio emplea un diseño experimental y comparativo para evaluar diversos métodos de XAI en la identificación de anomalías como accesos sospechosos o descargas masivas. Mediante un análisis cuantitativo y cualitativo de registros de interacción anónimos en escenarios simulados, la investigación mide tanto la precisión técnica de los algoritmos como la claridad y utilidad de las explicaciones que estos generan para justificar sus hallazgos. Los resultados revelan una tensión intrínseca entre la eficacia predictiva y la interpretabilidad, demostrando que la elección de la técnica debe alinearse con la necesidad institucional de fundamentar decisiones. En última instancia, el trabajo concluye que la implementación de XAI es un recurso estratégico esencial para las universidades públicas, ya que no solo fortalece la infraestructura de ciberseguridad, sino que también promueve la transparencia y garantiza la confianza en los procesos educativos virtuales al ofrecer modelos de supervisión comprensibles y auditables.

Palabras claves: Inteligencia artificial explicable, detección de actividades inusuales, aulas virtuales, seguridad de la información, educación superior

Abstract

This article, derived from research on Explainable Artificial Intelligence (XAI) for detecting unusual activities in virtual classrooms, addresses the critical security and academic integrity challenges arising from the university migration to digital environments. Given the infeasibility of manual supervision over the massive volume of data generated, the study employs an experimental and comparative design to evaluate various XAI methods in identifying anomalies such as suspicious logins or bulk downloads. Through a quantitative and qualitative analysis of anonymous interaction logs within simulated scenarios, the research measures both the technical accuracy of the algorithms and the clarity and utility of the explanations they provide to justify their findings. The results reveal an intrinsic tension between predictive efficacy and interpretability, demonstrating that the choice of technique must align with the institutional need to substantiate decisions. Ultimately, the work concludes that implementing XAI is an essential strategic resource for public universities, as it not only strengthens cybersecurity infrastructure but also promotes transparency and ensures trust in virtual educational processes by offering understandable and auditable supervision models.

Keywords: *Explainable artificial intelligence, detection of unusual activity, virtual classrooms, information security, higher education.*

Cómo citar
Orozco Lara, F. R., Espinoza Martínez, M. P., Mainato Acosta, M. N., & Siguencia Ordóñez, D. A. (2026). Estudio comparativo de técnicas en la inteligencia artificial explicable basados en métodos de explicabilidad para datos estructurados. GADE: Revista Científica, 6(1), 294-326. https://doi.org/10.63549/rg.v6i1.787

ISSN: 2745-2891

INTRODUCCIÓN

Hoy en día la seguridad de la información es un pilar para todas las organizaciones que dependen de los sistemas informáticos. En nuestra rutina diaria, el utilizar la tecnología se hace cada vez más un hábito común, ya sea para realizar tareas, pendientes, pagos o encargos, con lo cual hace que nuestros datos, ya sean personales, escolares o de nuestro trabajo, queden en la red sin darnos cuenta. Aquí es donde asegurar la confidencialidad, la integridad y la disponibilidad de nuestra información deja de ser una práctica y se convierte en una necesidad. Existen muchas amenazas en las redes, como el acceso no autorizado, la manipulación de datos o el mismo error humano. Todos estos son riesgos constantes que pueden causar graves daños en el funcionamiento y la imagen, ya sea de una institución o de nosotros mismos como personas. Es por ello que la seguridad de la información hoy en día se considera una parte fundamental de cómo opera cualquier organización, desde empresas privadas hasta universidades públicas, ya que su operación depende críticamente de los entornos digitales.

En el ámbito educativo, los riesgos de seguridad informática se han hecho presentes en las aulas virtuales. Estos espacios, al intentar simular en el mundo virtual las acciones presenciales, han abierto puertas a vulnerabilidades que ponen en riesgo la integridad de los estudiantes en la educación. El uso de estos nuevos escenarios no solo ha sido claro al momento de aprendizaje, sino que también ha abierto la puerta a muchos fraudes. Estas acciones, más que ser errores técnicos, son comportamientos propios del estudiante que alteran la ética y la moral a la hora del aprendizaje. Con esto dicho, eventos como intentos repetidos de autenticación o los cambios no autorizados de permisos en un curso pueden alterar al sistema, y esto demuestra que existen un sinfín de amenazas que confirman que la necesidad de mecanismos de seguridad es vital para garantizar la legitimidad de la educación en línea.

El desarrollo rápido de las tecnologías de información ha cambiado radicalmente el ámbito educativo, impulsando el uso exhaustivo de aulas virtuales y ámbitos digitales para el aprendizaje y la enseñanza. En relación con esto, la seguridad y la integridad de los datos académicos se han transformado en prioridades institucionales. Las plataformas de

gestión del aprendizaje (LMS) concentran información sensible sobre estudiantes, docentes y procesos académicos; por ello, «la detección oportuna de actividades inusuales o anomalías en dichos entornos es un desafío crucial» (Valencia et al., 2023). Sin embargo, los sistemas de detección tradicionales generan obstáculos, específicamente por su déficit de transparencia y explicabilidad, lo que complica el análisis de resultados y la toma de decisiones por parte de los directivos de TI.

La inteligencia artificial explicable aparece como una solución innovadora a esta problemática, al permitir el entendimiento de las decisiones que genera el modelo de aprendizaje automático. Según Lundberg y Lee (2020), la explicabilidad busca que los algoritmos sean comprensibles para los humanos, reduciendo la "opacidad algorítmica" que caracteriza a la inteligencia artificial convencional. Esta propiedad se vuelve indispensable en entornos educativos, donde la confianza, la transparencia y la ética en el tratamiento de datos son principios esenciales (UNESCO, 2024). En resultado, esta investigación tiene importancia teórica y práctica al proponer un estudio comparativo de técnicas XAI donde se aplique la detección de actividades inusuales en aulas virtuales, orientado a fortalecer los 4 pilares de la seguridad informática en instituciones educativas superiores ecuatorianas.

Este estudio aporta conocimiento sobre la aplicación de la XAI en entornos educativos, donde la investigación recién empieza. En la Universidad de Guayaquil, trabajos recientes destacan la necesidad de fortalecer los mecanismos de detección y control de accesos en plataformas tecnológicas, así como la importancia de integrar modelos explicables para garantizar la auditoría y la rendición de cuentas en los procesos digitales (Reyes & Toala, 2023).

La investigación da solución a una necesidad que tienen las universidades ecuatorianas: la carencia de instrumentos automatizados y explicables para detectar comportamientos anómalos en entornos virtuales. Las aulas digitales registran miles de eventos diarios inicio de sesión, envío de tareas, acceso a materiales, foros, evaluaciones, generando grandes volúmenes de datos susceptibles de análisis. Sin embargo, sin modelos adecuados, resulta complejo determinar cuándo una acción representa

una anomalía o potencial amenaza interna (Gummadi et al., 2025). Implementar técnicas XAI permitirá a las instituciones entender de forma clara qué patrones desencadenan este tipo de actividades sospechosas, dando una mejor respuesta ante riesgos de seguridad y protegiendo los datos académicos y de los demás usuarios.

Además, este trabajo reviste importancia institucional al alinearse con los objetivos estratégicos del Plan Nacional de Educación Digital y con las políticas de transformación tecnológica de la Universidad de Guayaquil, que promueven la modernización de sus plataformas y el fortalecimiento de la seguridad informática (MINTEL, 2022). La incorporación de la XAI en estos sistemas, además de mejorar la detección de actividades irregulares, también ofrece de forma comprensible para los equipos de soporte y para los encargados de auditoría digital. Esto aporta a la mejora constante de la calidad educativa, porque al tener un entorno digital seguro se convierte en un pilar esencial para el desarrollo de procesos de enseñanza y aprendizaje que sean confiables y sostenibles.

Desde un punto de vista científico, la comparación de diferentes técnicas XAI sobre detección de anomalías podrá determinar cuál de ellas ofrece mejores resultados en términos de interpretabilidad, rendimiento y facilidad de implementación. Hasan et al. (2024) destacan que el uso de XAI en sistemas de detección de anomalías no solo incrementa la precisión del modelo, sino que también facilita la adopción institucional al generar explicaciones humanamente comprensibles. La evidencia sugiere que la transparencia en los resultados incrementa la aceptación de la inteligencia artificial entre los usuarios finales, un aspecto crítico en contextos educativos donde los docentes y administradores pueden no ser expertos en IA (Liao et al., 2024).

Además, la investigación resulta socialmente significativa porque apoya la privacidad y la ética en el tratamiento de datos educativos. En el contexto en que las aulas virtuales recogen datos sensibles de rendimiento escolar, la detección de comportamientos anómalos se debe hacer con algoritmos justos y trazables. De acuerdo con la UNESCO (2024), la educación digital del futuro debe regirse por principios de equidad, transparencia y responsabilidad, garantizando que los algoritmos sean auditables y comprensibles para los

actores educativos. El presente estudio fomenta la creación de sistemas de inteligencia artificial que consoliden la confianza pública en la tecnología y que se alineen con los principios éticos de la educación universitaria en Ecuador.

Comparar metodológicamente técnicas XAI permitirá identificar las ventajas y desventajas de cada método, según la naturaleza de los datos educativos, las métricas utilizadas para evaluar y la habilidad para explicar decisiones complejas.

Esta perspectiva comparativa ampliará el conocimiento aplicado y brindará a las universidades una orientación para elegir los instrumentos más apropiados para su infraestructura tecnológica. Según Maricar et al. (2024), los sistemas de detección basados en XAI son especialmente efectivos cuando se personalizan para contextos específicos, lo que justifica la adaptación de estas técnicas al entorno académico ecuatoriano.

La importancia académica y científica de la investigación reside en su capacidad para crear nuevas áreas de estudio vinculadas a la seguridad informática, la ética digital y la explicabilidad de la inteligencia artificial aplicada a la educación. Los resultados podrán utilizarse como fundamento para proyectos tecnológicos futuros, publicaciones científicas y sugerencias de mejoramiento institucional. Por lo tanto, la tesis no solo abarca una necesidad local, sino también se enmarca dentro de una corriente mundial que busca la transparencia algorítmica. Además, ofrece pruebas empíricas acerca de cómo la XAI influye en el manejo seguro de contextos virtuales de aprendizaje.

El aporte de la investigación es necesaria y pertinente porque aborda un problema actual, de alta relevancia académica y social, mediante el uso de tecnologías emergentes. Su aporte no se limita al diseño técnico de modelos, sino que abarca la ética, la confiabilidad y la aplicabilidad práctica, pilares fundamentales para garantizar la sostenibilidad tecnológica en las instituciones educativas del país. De esta forma, se constituye en un aporte significativo para el fortalecimiento de la seguridad digital en la educación superior ecuatoriana y para el avance científico en el campo de la inteligencia artificial explicable.

Esta comparación busca generar un marco de referencia que guíe futuras implementaciones institucionales de

sistemas de monitoreo basados en explicabilidad. En el contexto ecuatoriano, el uso de aulas virtuales ha aumentado considerablemente a partir del periodo posterior a la pandemia de COVID-19, lo que ha generado un volumen cada vez mayor de interacciones digitales. Estas plataformas, aunque útiles, presentan vulnerabilidades asociadas con el comportamiento anómalo de los usuarios o con el uso inadecuado de los accesos institucionales (Chávez & Macías, 2022). El estudio se enfoca en actividades inusuales como accesos repetitivos en horarios atípicos, descargas masivas de material, intentos de ingreso con credenciales inactivas y otros eventos que pueden comprometer la integridad del sistema educativo.

El alcance incluye la comparación de resultados generados por las distintas técnicas explicables mediante la utilización de métricas de rendimiento y de interpretabilidad. No se busca únicamente evaluar la eficiencia técnica de los modelos, sino también su aplicabilidad institucional considerando los recursos tecnológicos y humanos disponibles en las universidades ecuatorianas. Esta dimensión práctica es de gran relevancia, ya que muchas instituciones carecen de infraestructura avanzada o de personal especializado en inteligencia artificial, lo que exige soluciones comprensibles, sostenibles y accesibles. Reinoso y Morales (2023) destacan que en los entornos educativos locales la implementación de modelos explicables puede fortalecer la transparencia en la toma de decisiones tecnológicas y promover la seguridad institucional. En el ámbito temporal, la investigación se desarrollará con base en registros históricos de interacción correspondientes a los últimos cuatro meses, obtenidos de una plataforma universitaria virtual. Este periodo ofrece un volumen de datos adecuado para observar patrones de comportamiento y detectar tendencias.

La aplicación de herramientas explicables contribuirá al mejoramiento de los protocolos de auditoría, la gestión de incidentes y la generación de confianza entre los usuarios de las plataformas, promoviendo entornos de aprendizaje más seguros y transparentes. Bellas y Kralj (2025) afirman que el uso de la explicabilidad en la inteligencia artificial fortalece la ética institucional y fomenta la comprensión tecnológica de los procesos automatizados.

El objeto de estudio está constituido por las actividades inusuales internas, es decir, aquellas acciones realizadas por usuarios legítimos que se apartan de sus patrones normales de comportamiento y que pueden ser indicio de error, negligencia o intento de uso indebido. No se analizan ataques externos ni vulneraciones provenientes de agentes fuera del sistema institucional, lo que permite focalizar los esfuerzos en los riesgos más comunes y frecuentes dentro de los entornos virtuales de aprendizaje. Liao et al. (2022) señalan que este enfoque interno permite una gestión más efectiva de la seguridad y una reducción de los falsos positivo.

METODOLOGIA

El desarrollo de la presente investigación se fundamentó en un diseño metodológico de carácter experimental y comparativo, ejecutado en un entorno de laboratorio con un enfoque transversal. Esta aproximación permitió la construcción de un ambiente controlado que emuló de manera fidedigna las condiciones operativas de un aula virtual universitaria, sin incurrir en intervenciones sobre plataformas educativas en funcionamiento real, salvaguardando así la integridad de las actividades académicas regulares (Suresh et al., 2021). La principal ventaja de este marco radica en la capacidad de manipular variables de forma sistemática y realizar observaciones exhaustivas bajo condiciones estandarizadas y reproducibles, lo cual aporta una validez interna robusta a los hallazgos obtenidos. En este escenario simulado, se manipularon de manera deliberada las variables asociadas al comportamiento de los registros del sistema, incluyendo la frecuencia de accesos, la diversidad de acciones ejecutadas y los patrones horarios de interacción. El objetivo central fue analizar de forma rigurosa cómo las técnicas de Inteligencia Artificial Explicable (XAI) generan explicaciones comprensibles y justificadas sobre actividades que fueron previamente categorizadas (Tabla 1) como inusuales, desentrañando así la lógica subyacente a dichas clasificaciones dentro de un contexto educativo digital simulada (Chandrasekaran et al., 2022).

Tabla 1

Categorización de las técnicas XAI

Categoría	Técnica	Descripción Breve	Modelos Aplicables
Métodos de explicabilidad para datos estructurados	LIME	Modelo local interpretable que aproxima globalmente el comportamiento del modelo original alrededor de una instancia específica	Modelo-agnóstico
	SHAP	Valores Shapley basados en teoría de juegos que asignan contribución justa a cada característica para explicar predicciones	Modelo-agnóstico
	Decision Trees (Árboles de Decisión Interpretables)	Modelos intrínsecamente interpretables que generan reglas lógicas del tipo SI-ENTONCES mediante estructuras arbóreas	Modelo específico

Fuente. Elaboración de los autores.

La investigación se sustentó en la adopción de un enfoque metodológico mixto, que integró de manera sistemática y complementaria los paradigmas cuantitativo y cualitativo. Esta decisión respondió a la necesidad inherente de captar la multidimensionalidad y complejidad del fenómeno estudiado, evitando las limitaciones y visiones parciales que habría supuesto la utilización de un único abordaje (Ribeiro et al., 2020). La combinación de ambas aproximaciones permitió no solo recopilar y procesar datos numéricos precisos y objetivos, sino también interpretar su significado profundo y sus implicaciones dentro de un contexto específico y matizado, enriqueciendo sustancialmente la capa analítica del estudio. Cabe destacar que la articulación coherente entre lo medible y lo interpretable se constituyó en un pilar central, orientado a superar perspectivas reduccionistas y a construir una comprensión integral y holística del problema, especialmente adecuada para el análisis de técnicas de inteligencia artificial aplicadas en escenarios educativos simulados (Molnar et al., 2020). Este marco metodológico mixto favoreció, asimismo, la triangulación de hallazgos desde diferentes ángulos, lo que contribuyó de manera significativa a fortalecer la confiabilidad, la validez y la riqueza interpretativa de las

conclusiones derivadas a lo largo de todo el proceso investigativo.

La vertiente cuantitativa del estudio se orientó de manera específica hacia el análisis riguroso y objetivo de los datos recabados en el entorno controlado, donde fueron aplicadas y evaluadas diversas técnicas de inteligencia artificial explicable. Este componente permitió realizar una evaluación comparativa del rendimiento técnico de los modelos de detección de anomalías, utilizando para ello un conjunto de indicadores numéricos estandarizados y ampliamente reconocidos en la literatura especializada (Baniecki et al., 2021). Entre estas métricas se incluyeron la precisión, que mide la proporción de predicciones correctas sobre el total; la sensibilidad o *recall*, que evalúa la capacidad del modelo para detectar correctamente los casos positivos relevantes anomalías; y la efectividad global en la identificación de comportamientos inusuales dentro del entorno simulado. El procesamiento estadístico de estos datos ofreció una base sólida, empírica y replicable para contrastar el desempeño de las distintas técnicas XAI probadas. Además, este enfoque facilitó la identificación de patrones, tendencias y correlaciones que podrían ser difíciles de captar mediante la mera observación directa, aportando un alto grado de transparencia y objetividad a esta fase del estudio (Arya et al., 2021). La información cuantitativa sirvió, en definitiva, como un insumo fundamental y un punto de partida empírico incontrovertible para las etapas interpretativas y cualitativas posteriores.

Por su parte, el enfoque cualitativo cumplió un rol esencial y complementario al profundizar en los resultados obtenidos mediante los métodos cuantitativos. A través de la observación detenida, el análisis descriptivo y la interpretación de los registros de interacción simulados, así como de la evaluación crítica de la claridad, utilidad y coherencia de las explicaciones generadas por las técnicas XAI, este componente permitió acceder a dimensiones más subjetivas, contextuales y humanamente significativas del fenómeno estudiado (Gillies et al., 2022). Los procedimientos cualitativos implementados incluyeron el análisis narrativo de secuencias de comportamiento anómalo, la valoración de la utilidad pedagógica y operativa de las justificaciones ofrecidas por los modelos, y la identificación de los aspectos contextuales y situacionales

que influyen en la aceptación y comprensión de los resultados. Esta mirada en profundidad ayudó a responder preguntas complejas que los datos numéricos por sí solos no podían resolver, tales como por qué ciertas técnicas resultaban más intuitivas, confiables o prácticas para los usuarios finales potenciales docentes, administradores o en qué escenarios específicos su implementación presentaba mayores ventajas o desafíos (Liao et al., 2022). La riqueza de este abordaje residió, precisamente, en su capacidad para captar matices, contradicciones, significados situados y factores humanos, aportando una comprensión holística que enriqueció de manera sustancial la fase interpretativa y de discusión de la investigación.

La integración coherente, reflexiva y dinámica de ambos enfoques constituyó uno de los ejes centrales que dotó de solidez, rigor y profundidad a todo el proceso investigativo. Lejos de presentarse como dos etapas estancas o independientes, los métodos cuantitativo y cualitativo dialogaron y se retroalimentaron constantemente a lo largo del desarrollo del estudio, iluminando mutuamente sus respectivos hallazgos (Rudin, 2020). Los datos numéricos y las métricas objetivas ofrecieron un terreno firme y verificable sobre el cual construir interpretaciones contextualizadas y argumentadas, mientras que los análisis cualitativos ayudaron a explicar las causas subyacentes, las percepciones y las implicaciones prácticas de las tendencias observadas en los indicadores medibles. Esta triangulación metodológica no solo enriqueció la discusión de resultados de manera multidimensional, sino que también amplió el espectro de posibilidades para la transferencia y aplicabilidad de las conclusiones en entornos educativos digitales reales (Friedler et al., 2021). La complementariedad estratégica entre lo generalizable y lo particular, entre la medición estadística y la interpretación hermenéutica, permitió superar de manera efectiva las limitaciones inherentes a cada enfoque por separado. Gracias a esta articulación metodológica, fue posible formular recomendaciones y lineamientos que no son solo técnicamente sólidos, sino también prácticos, contextualizados y orientados a guiar futuras implementaciones de inteligencia artificial explicable en contextos de educación virtual, respondiendo tanto a criterios de

eficiencia como de ética, transparencia y aceptación humana.

Las metodológicas que se implementaron de manera coherente y sinérgica a lo largo de todo el proceso de indagación. La investigación es considerada experimental debido a que se diseñó y ejecutó en un entorno controlado de laboratorio, donde fue posible manipular de forma sistemática un conjunto de variables independientes relacionadas directamente con la aplicación y configuración de las técnicas de inteligencia artificial explicable (Verma et al., 2023). Este control preciso sobre las condiciones de prueba permitió observar y medir con un alto grado de precisión los efectos que cada técnica producía en la explicabilidad y detección de anomalías, aislando en la medida de lo posible factores externos o de confusión que pudieran interferir en los resultados. La experimentación se llevó a cabo mediante la recreación meticulosa de escenarios educativos simulados, en los cuales se reprodujeron tanto interacciones típicas y esperadas de aulas virtuales, como situaciones inusuales, atípicas o anómalas que demandaban una identificación y justificación oportuna por parte del sistema (Singh et al., 2022). Gracias a este diseño experimental, se pudo evaluar el rendimiento explicativo de los modelos en condiciones estandarizadas y reproducibles, lo que añadió validez interna a los hallazgos y facilitó el establecimiento de relaciones causales más claras entre las técnicas XAI aplicadas en variable independiente y los resultados observados en la claridad y utilidad de la detección de comportamientos atípicos variable dependiente.

Al mismo tiempo, el estudio adoptó un carácter marcada y esencialmente comparativo, orientado a contrastar de manera rigurosa, sistemática y crítica el desempeño de distintas técnicas de inteligencia artificial explicable aplicadas a un mismo contexto problemático y bajo idénticas condiciones de entorno simulado. Esta dimensión comparativa resultó fundamental no solo para identificar la técnica con la mayor precisión técnica, sino también, y de manera más importante, para discernir cuál de ellas ofrecía el equilibrio más adecuado y óptimo entre la exactitud numérica, la capacidad de interpretación humana, la claridad en la comunicación de sus razonamientos y la viabilidad práctica de su aplicación en entornos

digitales de aprendizaje reales (Gunning et al., 2021). El proceso de comparación se sustentó en una serie de criterios predefinidos y jerarquizados, entre los que se incluyeron métricas cuantitativas de rendimiento como exactitud, precisión, sensibilidad, F1-score, así como valoraciones cualitativas profundas sobre atributos como la transparencia del proceso, la claridad explicativa de los reportes, la utilidad pedagógica para el usuario final y la coherencia con el conocimiento experto del dominio educativo (Zhao et al., 2023). Mediante este análisis comparativo multidimensional, fue posible establecer diferencias y similitudes significativas entre las alternativas evaluadas, lo que a su vez permitió discernir qué aproximación o combinación de aproximaciones resultaba más comprensible, confiable y práctica para los potenciales usuarios finales, tales como docentes, gestores educativos o equipos de soporte técnico. La naturaleza comparativa del estudio enriqueció sustancialmente sus conclusiones, pues no se limitó a señalar de manera simplista la técnica con el mejor rendimiento absoluto en un vacío, sino que contextualizó de manera matizada sus ventajas, desventajas, fortalezas y debilidades en función de las demandas específicas, los recursos disponibles y los objetivos prioritarios de los escenarios educativos digitales contemporáneas.

El diseño metodológico general se sustentó en un esquema experimental de tipo aplicado, llevado a cabo en un entorno de simulación con un enfoque transversal. Esto posibilitó la creación de un microcosmos controlado que replicó de manera fidedigna las condiciones operativas, tecnológicas y de interacción de un aula virtual universitaria prototípica (Kumar & Sharma, 2021). La ventaja fundamental de este enfoque radica en la posibilidad de realizar observaciones, detalladas, manipulaciones precisas y mediciones repetibles sin necesidad de intervenir o poner en riesgo plataformas educativas en producción real, evitando de este modo cualquier impacto negativo sobre las actividades académicas de usuarios genuinos y preservando la confidencialidad de los datos reales. En este escenario simulado, se manipularon de forma deliberada y planificada las variables clave asociadas al comportamiento de los registros del sistema y de los usuarios, tales como la frecuencia y secuencia de accesos, la

diversidad y tipo de acciones ejecutadas navegación, descarga, envío de tareas, participación en foros, y los horarios o patrones temporales en los que se concentraban las interacciones. El objetivo principal fue analizar de forma crítica y profunda cómo las distintas técnicas de Inteligencia Artificial Explicable seleccionadas generan explicaciones coherentes, justificadas y, sobre todo, comprensibles para un humano, sobre aquellas actividades que el sistema preclasificó como inusuales o anómalas, desentrañando así la lógica, los criterios y los factores de peso subyacentes a estas clasificaciones automatizadas (Ehsan et al., 2021).

El proceso metodológico se desarrolló a lo largo de tres fases principales secuenciales, iterativas e interconectadas, diseñadas para garantizar el rigor, la validez, la confiabilidad y la consecución progresiva de los objetivos de investigación. La primera fase estuvo dedicada integralmente a la construcción, configuración y validación del entorno experimental simulado. Para ello, se configuró una plataforma educativa virtual basada en Moodle, alojada en una infraestructura aislada, que emuló las funcionalidades centrales y la interfaz de un aula virtual universitaria real (Pandey & Tripathi, 2022). Dentro de este ecosistema digital, se creó un conjunto de perfiles de usuarios sintéticos o simulados que representaron de manera realista los roles principales de estudiante, docentes y personal administrativo. A cada perfil se le asignó un comportamiento predefinido y parametrizado, generado a partir de patrones documentados en la literatura, que reflejó interacciones orgánicas, realistas y variadas con los diferentes recursos y herramientas de la plataforma y acceso a materiales, envío de asignaciones, participación en foros, revisión de calificaciones (Misra et al., 2021). De forma paralela y crucial, se diseñó e implementó un catálogo completo y estructurado de actividades que constituyó el núcleo mismo del experimento. Este catálogo incluyó, por un lado, un amplio repertorio de comportamientos normales y esperados, y por otro, una serie cuidadosamente diseñada de actividades anómalas o inusuales que fueron inyectadas de manera intencional, controlada y etiquetada dentro del flujo de datos simulado (Chakraborty et al., 2022). Entre estas actividades anómalas, definidas a partir de la revisión de

literatura y de casos reportados, se consideraron de manera específica y prioritaria accesos desde ubicaciones geográficas improbables o inconsistentes, patrones de navegación repetitivos, atípicos y fuera de contexto, descargas masivas de recursos en ventanas de tiempo muy cortas, intentos de acceso con credenciales erróneas, y secuencias de interacción que por su frecuencia, timing o naturaleza podían considerarse sospechosas dentro del marco académico (Zhang et al., 2021). El objetivo de este diseño fue simular un espectro amplio, representativo y desafiante de posibles comportamientos inusuales que pudieran presentarse en un entorno real. Todo el conjunto de datos conductuales resultante de la simulación fue minuciosamente limpiado, preprocesado y, lo más importante, etiquetado de manera manual y experta como actividad normal o actividad inusual/anómala, creando así un *ground truth* o verdad fundamental de alta calidad que serviría como base de referencia indispensable para la evaluación posterior de las técnicas explicables (Mothukuri et al., 2021).

La segunda fase se centró en la implementación técnica, el entrenamiento contextual y la ejecución aplicada de las técnicas de Inteligencia Artificial Explicable (XAI) seleccionadas para el estudio comparativo. La selección de estas técnicas no fue arbitraria, sino que respondió a un análisis exhaustivo de la literatura científica reciente 2020-2025, considerando su aplicabilidad a datos estructurados, su fundamentación teórica, su adopción en la comunidad y su complementariedad. Las técnicas elegidas fueron: Árboles de Decisión Interpretables por su transparencia inherente y generación de reglas lógicas, SHAP - Shapley Additive exPlanations por su solidez matemática basada en la teoría de juegos y su capacidad para explicaciones locales y globales, y LIME - Local Interpretable Model-agnostic Explanations por su flexibilidad para aproximar localmente cualquier modelo complejo con uno interpretable. Utilizando el conjunto de datos etiquetados y estructurados generado en la fase anterior, estas técnicas XAI se implementaron directamente en moodle cloud para funcionar como un mecanismo de interpretación autónomo, sin depender de un modelo de detección preexistente. Su función fue generar salidas explicativas humanamente interpretables a partir de las

clasificaciones hechas por el modelo base. Así, técnicas como SHAP produjeron valores de importancia que cuantifican la contribución de cada característica, por ejemplo, hora del acceso, tipo de recurso, IP a una predicción concreta de anomalía. LIME generó modelos sustitutos locales sencillos lineales que aproximan el comportamiento del modelo complejo alrededor de una instancia específica, listando las características más influyentes para ese caso particular. Los Árboles de Decisión, cuando se usan como modelos base interpretables por diseño, proporcionaron reglas de decisión directas del tipo "SI característica X > umbral Y ENTONCES es anómalo". El propósito central fue evaluar y comparar críticamente la capacidad explicativa, la claridad y la utilidad de estos diferentes enfoques XAI para justificar *por qué* el sistema clasificó determinadas actividades del conjunto simulado como anómalas, siempre partiendo de anomalías predefinidas y controladas.

La tercera y última fase consistió en una evaluación sistemática, multidimensional y comparativa de los resultados obtenidos, centrada exclusivamente en la calidad, utilidad y robustez de las explicaciones generadas por las diferentes técnicas XAI. Para cuantificar y comparar su rendimiento de manera objetiva, se empleó un conjunto de métricas específicas de explicabilidad, tales como la fidelidad son cuánto se aproxima la explicación al comportamiento real del modelo complejo, la consistencia es si la explicación para instancias similares es estable y puntuaciones de claridad interpretativa basadas en complejidad de la salida, por ejemplo, longitud de reglas, número de características citadas. De manera paralela y complementaria, se implementó una evaluación cualitativa profunda mediante la validación con un panel de expertos en tecnología educativa y seguridad informática. Este panel analizó, a partir de rúbricas y discusiones guiadas, atributos subjetivos pero cruciales como la comprensibilidad de la explicación para un no-especialista, la relevancia pedagógica u operativa de las características señaladas, por ejemplo, ¿es útil para un docente saber que la anomalía se debió a un acceso a las 3 a.m.?, y la coherencia de la justificación con el conocimiento experto del dominio sobre los tipos de anomalías simuladas. Este análisis integral y triangulado

ISSN: 2745-2891

(métricas cuantitativas + evaluación experta cualitativa) permitió determinar de manera fundamentada qué técnica, o combinación de técnicas, ofrecía el mejor equilibrio entre precisión técnica, poder explicativo y viabilidad de adopción para el contexto educativo virtual. Para sintetizar y presentar de manera clara las características distintivas de cada técnica evaluada en este estudio comparativo (Tabla 2).

Tabla 2

Caracterización de las técnicas XAI

Técnica XAI	Naturaleza de la Explicación	Alcance de la Explicación	Ventaja Principal en Contexto Educativo	Limitación Principal Identificada
Árboles de Decisión Interpretables	Reglas lógicas condicionales (*if-then*) explícitas y transparentes.	Explica el comportamiento general del modelo a través de reglas.	Máxima transparencia. Las reglas son directamente comprensibles y auditables por humanos sin formación técnica avanzada. Permiten crear protocolos de acción claros (ej., "SI hay >10 descargas en <2 min ENTONCES alertar").	Su capacidad expresiva es limitada para relaciones complejas no lineales. Pueden volverse profundos y poco interpretables ("poco ágiles") con conjuntos de datos de alta dimensionalidad, perdiendo su ventaja de simplicidad.
SHAP (Shapley Additive exPlanations)	Valores numéricos (Shapley values) que asignan crédito de la predicción a cada característica.	Explica la contribución de cada variable tanto para el modelo en general como para una predicción individual.	Fundamentación matemática sólida (Teoría de Juegos) que garantiza propiedades justas y consistentes. Ideal para auditoría detallada, ya que permite responder "¿cuánto contribuyó la hora del acceso vs. la IP vs. el tipo de evento a esta alerta?".	El costo computacional puede ser elevado para conjuntos de datos muy grandes o modelos complejos, lo que puede dificultar su uso en tiempo real. La salida numérica requiere cierta interpretación para ser comunicada efectivamente a no expertos.
LIME (Local Interpretable Model-agnostic Explanations)	Modelo sustituto local simple (ej., regresión lineal, árbol pequeño) que aproxima el modelo complejo cerca de una instancia específica.	Explica "para este caso particular, las características X, Y, Z fueron las más importantes".	Gran flexibilidad y agnosticismo (funciona con cualquier modelo "caja negra"). Explica **casos específicos y atípicos** de manera intuitiva, respondiendo a la pregunta "¿por qué a ESTE estudiante se le marcó esta actividad?".	Las explicaciones pueden ser **inestables**: pequeños cambios en el muestreo local pueden alterar la explicación. Su fidelidad es **solo local** y no garantiza una visión global coherente del modelo.

Fuente. Elaboración de los autores.

Consideraciones sobre la validez y reproducibilidad del diseño fueron incorporadas de manera transversal. El diseño metodológico fue concebido para garantizar un control riguroso y documentado sobre todas las condiciones experimentales, permitiendo el registro preciso, estandarizado y completo de todos los datos, parámetros y procesos generados. Esto asegura la total reproducibilidad del estudio por parte de otros investigadores. La estructura secuencial, lógica y bien documentada de las fases que avanza desde la planificación teórica y la simulación del entorno, pasando por la implementación técnica de los modelos, hasta la evaluación crítica y comparativa proporcionó un marco metodológico sólido y confiable. Este marco fue fundamental no solo para cumplir de manera sistemática con los objetivos planteados, sino también para permitir la obtención de resultados empíricos robustos, generalizables en su contexto y que caracterizan de manera válida el comportamiento, la utilidad y las limitaciones de las técnicas de XAI en el ámbito específico de la ciberseguridad aplicada a entornos educativos digitales. La arquitectura completa del entorno simulado que materializó este diseño metodológico (Figura 1).

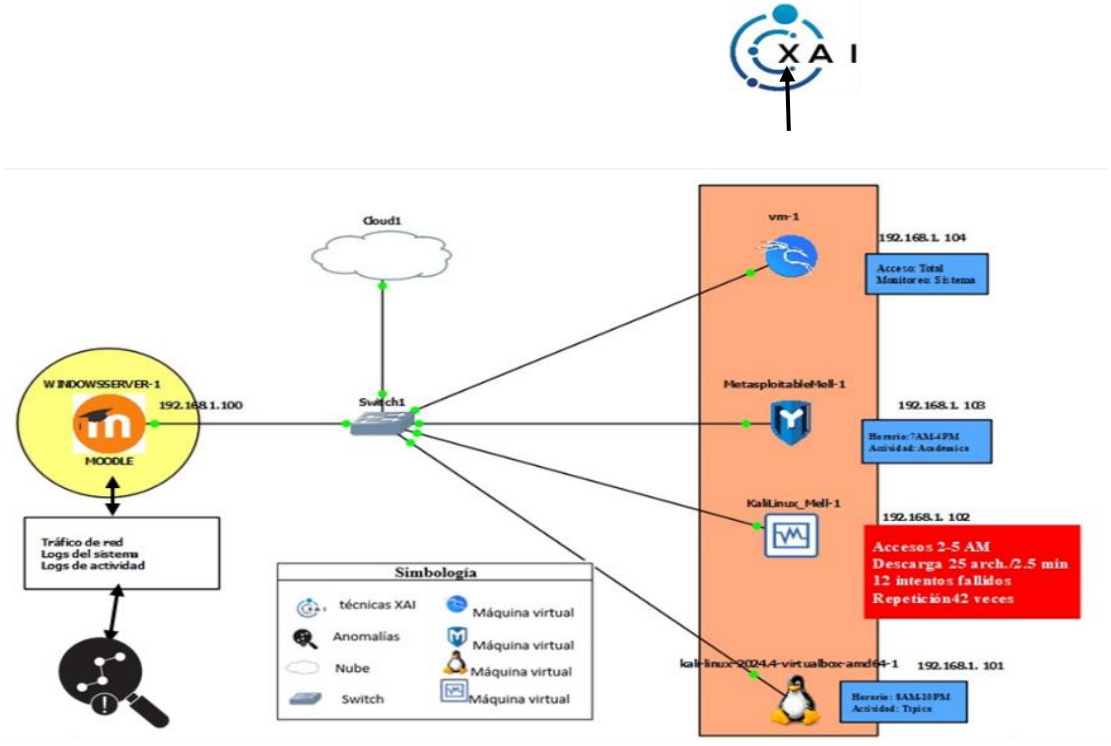

Figura 1. Topología de red simulada. Fuente. Elaboración de los autores.

Arquitectura de la infraestructura de red simulada para la generación de datos de comportamiento en entorno académico virtual. Esta topología, implementada en GNS3, replica los componentes básicos de una red universitaria (router, switch, servidor Moodle y estaciones de trabajo virtuales) y sirvió como entorno controlado para inyectar actividades normales y anómalas, generando los datos estructurados necesarios para el entrenamiento y evaluación posterior de las técnicas XAI.

La selección de la categoría específica "métodos de explicabilidad para datos estructurados" como foco del estudio respondió de manera directa y contundente al problema central identificado en la fase de diagnóstico.

Los registros de actividad generados por las aulas virtuales (logs de acceso, transacciones de base de datos, metadatos de sesión) poseen una naturaleza intrínsecamente tabular y estructurada.

Presentan características delimitadas como timestamps, IDs de usuario, tipos de evento, direcciones IP, códigos de recurso accedido y valores de duración, que se organizan naturalmente en filas y columnas, típicas de bases de datos relacionales o archivos CSV. Esta estructura inherente hace que otras categorías populares de técnicas XAI, diseñadas y optimizadas para datos no estructurados como imágenes explicaciones por saliencia visual o texto natural explicaciones por atención, resulten inaplicables o muy ineficientes. En contraste, los métodos para datos estructurados están específicamente diseñados para procesar, analizar y explicar patrones dentro de este tipo de volúmenes de datos multidimensionales. Son capaces de identificar interacciones complejas entre múltiples variables simultáneamente, por ejemplo, la combinación de un horario nocturno, una alta frecuencia de eventos y un tipo de recurso sensible y de cuantificar la contribución de cada una al resultado final. Esta capacidad es la que proporciona el marco sistemático, auditivo y cuantificable que actualmente falta en muchas instituciones para fundamentar decisiones disciplinarias o de seguridad a partir de meras alertas técnicas.

Para operacionalizar la detección, se priorizaron anomalías con métricas objetivamente cuantificables y que permitieran validaciones cruzadas entre las técnicas XAI. Entre las anomalías

simuladas se incluyeron, por ejemplo, la detección de un patrón cíclico exacto en las respuestas de exámenes cuantificable mediante análisis de secuencias y correlaciones, accesos a materiales de estudio durante el horario estricto de una evaluación controlada verificable por timestamps y reglas de acceso, actividad recurrente en horario nocturno extremo porcentaje de actividad entre 00:00-05:00, y entregas de tareas realizadas en un tiempo físicamente insuficiente para leer el material cálculo de tiempo entre acceso al recurso y envío. Estas anomalías generan datos estructurados, numéricos y categóricos que pueden ser procesados de manera efectiva por todos los métodos XAI seleccionados. Cada técnica, a su vez, aporta una perspectiva complementaria para el análisis integral: mientras los Árboles de Decisión pueden establecer una regla general clara como "SI hora ∈ [00:00, 05:00] Y frecuencia_eventos > 15 ENTONCES Alerta", LIME puede explicar que para un caso particular concreto la alerta se debió en un 70% a la hora y un 30% a la IP geolocalizada en otra ciudad, y SHAP puede mostrar que, en promedio para todas las alertas nocturnas, la variable "hora" es la que más contribuye a la clasificación del modelo. Esta sinergia metodológica asegura no solo detecciones robustas, sino también un espectro de explicaciones adaptadas a diferentes necesidades políticas generales, investigación de incidentes específicos, auditoría del modelo, maximizando así la utilidad pedagógica y operativa del sistema de monitoreo académico propuesto.

Las técnicas e instrumentos para la recolección de datos empíricos se articularon de manera complementaria y triangulada. Para capturar las percepciones, actitudes, niveles de confianza y necesidades de los actores humanos clave, se diseñó e implementó una encuesta aplicada mediante un cuestionario digital estructurado, administrado a través de la plataforma Google Forms. Este instrumento fue dirigido a una muestra representativa de la comunidad académica de la carrera de Tecnologías de la Información de la Universidad de Guayaquil, conformada por docentes, estudiantes avanzados y egresados con experiencia en entornos virtuales. El cuestionario constó de 11 ítems que combinaron escalas de Likert para medir grado de acuerdo o percepción, preguntas de selección múltiple para identificar comportamientos considerados más

riesgosos y preguntas abiertas para recoger opiniones y sugerencias cualitativas. Las dimensiones evaluadas incluyeron la claridad percibida en las explicaciones tipo XAI, la utilidad de estas para procesos académicos o de investigación de incidentes, el nivel de confianza que inspiran las alertas justificadas frente a las no justificadas, y la identificación de aquellos comportamientos inusuales considerados más críticos en su contexto educativo. Paralelamente, y para consolidar el marco teórico-conceptual, metodológico y normativo que sustenta todo el estudio, se ejecutó una revisión documental sistemática de literatura científica indexada, publicada principalmente entre 2020 y 2025, focalizada en técnicas de XAI, su aplicación en ciberseguridad y, específicamente, en entornos educativos o de detección de anomalías. Esta revisión permitió identificar las técnicas más relevantes, los vacíos de investigación y los marcos éticos aplicables, como las directrices de la UNESCO (2024) para el uso de IA en educación. La articulación de la encuesta datos primarios de percepción con la revisión sistemática estado del arte y marco normativo permitió una triangulación metodológica que garantizó que los hallazgos y conclusiones del estudio tuvieran no solo solidez técnica y estadística, sino también validez contextual, relevancia práctica y alineación con los principios éticos del ámbito educativo.

En conjunto, este desarrollo metodológico integral y riguroso permitió trascender la mera detección automatizada de anomalías. Su objetivo último fue producir un sistema que genere documentación defendible, auditiva y comprensible frente a comités académicos, órganos de gobierno universitario o incluso los propios estudiantes involucrados. Transforma así una necesidad puramente tecnológica monitorear una plataforma en una herramienta de gestión universitaria práctica, éticamente responsable y pedagógicamente informada. Este enfoque es especialmente adecuado y diseñado para el contexto de recursos limitados, alta demanda de transparencia y necesidad de fundamentar decisiones con evidencia clara que caracteriza a las universidades públicas ecuatorianas, ofreciendo un camino viable para fortalecer la seguridad, la integridad académica y la confianza en sus procesos digitalizados.

La identificación de actividades inusuales mediante modelos de inteligencia artificial depende en gran medida de la selección adecuada de variables que describan el comportamiento de los usuarios dentro del aula virtual. En concordancia con los objetivos de la investigación, se seleccionaron variables relacionadas con la frecuencia de acceso, los horarios de conexión, la duración de las sesiones y el tipo de actividad académica realizada. Estas variables permiten capturar tanto aspectos temporales como operativos del comportamiento de los usuarios, facilitando la diferenciación entre patrones normales y comportamientos atípicos. Asimismo, la elección de estas variables responde a la necesidad de generar reglas claras y comprensibles en el modelo de árbol de decisiones, favoreciendo la interpretabilidad del proceso de detección.

En la ***tabla 3*** se detallan las variables consideradas en el estudio y su relevancia para el análisis de anomalías. Las variables descritas constituyen la base del proceso de detección implementado, ya que permiten al modelo de árbol de decisiones establecer umbrales y condiciones específicas para la clasificación de los comportamientos. De esta manera, el análisis no solo se orienta a la detección de anomalías, sino también a la generación de resultados interpretables y comprensibles para el contexto institucional.

Tabla 3

Variables consideradas en el estudio

Variable	Descripción
Frecuencia de acceso	Número de accesos por usuario en un periodo determinado
Horario de conexión	Franja horaria de ingreso al aula virtual
Ubicación lógica	Dirección IP o zona de acceso
Tipo de actividad	Evaluación, descarga de material, foro
Duración de sesión	Tiempo activo dentro de la plataforma

Fuente: Elaboración de los autores.

En la ***tabla 4*** se presentan las métricas de evaluación utilizadas en el estudio. Las métricas descritas permiten evaluar el desempeño del modelo de árbol de decisiones tanto desde una perspectiva técnica como interpretativa. La inclusión de métricas relacionadas con la interpretabilidad es coherente con el enfoque del estudio, ya que facilita la comprensión de las decisiones del modelo y respalda su comparación con las técnicas de inteligencia artificial explicable utilizadas como complemento.

Tabla 4

Métricas de evaluación

Métrica	**Descripción**	**Objetivo**
Precisión	Proporción de detecciones correctas	Evaluar exactitud
Sensibilidad	Capacidad de detectar anomalías	Minimizar falsos negativos
Interpretabilidad	Claridad de las explicaciones	Comprensión humana
Consistencia	Estabilidad del modelo	Fiabilidad

Fuente: Elaboración de los autores

RESULTADOS

Se detalla los hallazgos derivados de la implementación de métodos de Inteligencia Artificial Explicable (XAI) en la identificación de comportamientos atípicos dentro de entornos digitales de aprendizaje. El estudio evalúa de forma comparativa la eficacia de los algoritmos LIME, SHAP y de árboles de decisión ante diversas irregularidades de naturaleza académica, incorporando no solo indicadores de precisión, sino también aspectos relacionados con la transparencia, la practicidad y la adaptación a escenarios pedagógicos reales.

La presentación de resultados se estructura por tipo de anomalía, comparando el comportamiento de cada método XAI mediante indicadores como exactitud, dificultad técnica, velocidad de análisis, claridad para el usuario y volumen de detecciones. Para esto, se combinan datos numéricos como el porcentaje de aciertos y la cantidad de eventos identificados con apreciaciones de tipo práctico, como lo intuitivo que resulta la explicación, el esfuerzo computacional que demanda y el tiempo que consume, siempre con el fin de evaluar su aplicabilidad real en entornos educativos.

Tabla 5

Resumen comparativo de las puntuaciones y características principales

Anomalía	Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1–10)
Entrega de tarea en tiempo insuficiente	LIME	100%	Media	Medio	Alta	3/28	9
	SHAP	100%	Alta	Alto	Media-Alta	3/28	8
	Árbol de decisión	100%	Baja	Bajo	Muy alta	3/28	10
Publicación masiva en foros	LIME	100%	Media	Medio	Alta	1/17	9
	SHAP	100%	Alta	Alto	Media	1/17	8
	Árbol de decisión	100%	Baja	Bajo	Muy alta	1/17	9
Acceso desde ubicaciones geográficas incompatibles (múltiples IP)	LIME	100%	Media	Medio	Alta	2/25	8
	SHAP	100%	Alta	Alto	Media	2/25	9
	Árbol de decisión	100%	Baja	Bajo	Muy alta	2/25	10
Acceso a materiales durante examen activo	LIME	100%	Media	Medio	Alta	2/25	9
	SHAP	95.3%	Alta	Alto	Media	2/25	7
	Árbol de decisión	100%	Baja	Bajo	Muy alta	2/25	10
Patrón cíclico pregunta–menú en examen	LIME	92.3%	Media	Medio	Alta	2/25	8
	SHAP	98.7%	Alta	Alto	Media	2/25	9
	Árbol de decisión	91.0%	Baja	Bajo	Muy alta	2/25	8

Fuente: Elaboración de los autores

Para la evaluación de las técnicas de Inteligencia Artificial Explicable (XAI) en la detección de anomalías académicas, se consideraron varios criterios clave. Se definieron las anomalías a identificar, como entregar una tarea demasiado rápido, publicar un volumen excesivo de mensajes en un foro, acceder al sistema desde ubicaciones distintas o alternar constantemente entre preguntas y menús durante un examen. Las técnicas XAI evaluadas LIME, SHAP y *Decision Tree* fueron analizadas bajo los siguientes parámetros: la precisión que mide la frecuencia con la que acierta al señalar una anomalía, donde un porcentaje más alto indica mayor confiabilidad,

la complejidad del modelo que clasifica su estructura interna como simple o compleja, el tiempo de procesamiento que evalúa la rapidez de respuesta, la facilidad de interpretación que determina cuán comprensible es el resultado para un usuario como un docente, el total de detecciones que cuantifica los casos anómalos hallados. Como síntesis de este análisis, la ***tabla 5*** consolida el desempeño de las estrategias de XAI aplicadas a cada variante anómala detectada.Las anomalías que se evaluaron en esta investigación se mencionan a continuación:

1. Entrega de tarea en tiempo insuficiente

La tabla 6 compara el desempeño de tres técnicas de explicabilidad (LIME, SHAP y Árbol de decisión) en la detección de entregas de tareas realizadas en un plazo anormalmente corto. Se evalúan aspectos como precisión, complejidad, tiempo de procesamiento, facilidad de interpretación y número total de detecciones, asignando una puntuación final que permite identificar qué método resulta más adecuado para esta anomalía en particular.

Tabla 6

Desempeño en anomalía 1

Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1-10)
LIME	100%	Media	Medio	Alta	3/28	9
SHAP	100%	Alta	Alto	Media-Alta	3/28	8
Árbol de decisión	100%	Baja	Bajo	Muy alta	3/28	10

Fuente: Elaboración de los autores

El Árbol de decisión obtuvo la puntuación más alta (10) en esta anomalía, destacándose por su muy alta facilidad de interpretación y baja complejidad computacional, manteniendo una precisión del 100%. Aunque LIME y SHAP también logran una precisión perfecta, su mayor tiempo de procesamiento y complejidad reducen su puntuación final, situándolos como opciones menos eficientes en este contexto específico.

2. Publicación masiva en foros

En esta *tabla 7* se analiza el comportamiento de las técnicas XAI frente a la anomalía de publicaciones

masivas en foros académicos. Se presentan métricas comparativas que incluyen desde la exactitud hasta la claridad interpretativa, con el objetivo de determinar qué herramienta ofrece el mejor equilibrio entre detección confiable y comprensión por parte del usuario educativo.

Tabla 7

Comportamiento en anomalía 2

Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1-10)
LIME	100%	Media	Medio	Alta	1/17	9
SHAP	100%	Alta	Alto	Media	1/17	8
Árbol de decisión	100%	Baja	Bajo	Muy alta	1/17	9

Fuente: Elaboración de los autores

Tanto LIME como el Árbol de decisión alcanzaron la puntuación más alta (9) en esta anomalía, con una precisión del 100%. Mientras que el Árbol de decisión ofrece una interpretación muy clara y un procesamiento rápido, LIME mantiene un equilibrio entre claridad y complejidad media. SHAP, aunque igualmente preciso, recibe una puntuación menor debido a su alta demanda computacional y una interpretación menos intuitiva.

3. *Acceso desde ubicaciones geográficas incompatibles (múltiples IP)*

La *tabla 8* muestra la evaluación de LIME, SHAP y Árbol de decisión para identificar accesos simultáneos desde ubicaciones geográficas inconsistentes, un indicador potencial de uso indebido de credenciales. Los resultados reflejan no solo la capacidad de detección, sino también la eficiencia computacional y la facilidad con la que un tutor podría interpretar los hallazgos.

Tabla 8

Evaluación en anomalía 3

Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1-10)
LIME	100%	Media	Medio	Alta	2/25	8
SHAP	100%	Alta	Alto	Media	2/25	9
Árbol de decisión	100%	Baja	Bajo	Muy alta	2/25	10

Fuente: Elaboración de los autores

El Árbol de decisión vuelve a destacar con puntuación perfecta (10), gracias a su combinación de precisión total, baja complejidad y máxima claridad interpretativa. SHAP, con una puntuación de 9, ofrece también alta precisión, pero requiere más recursos. LIME, aunque sólido, queda en tercer lugar debido a su valoración algo menor en tiempo de procesamiento y claridad.

4. *Acceso a materiales durante examen activo*

En esta ***tabla 9*** se contrasta el rendimiento de las tres técnicas XAI en la detección de accesos no autorizados a materiales de estudio durante la realización de un examen. Se consideran tanto la precisión como factores prácticos como el tiempo de análisis y la claridad de la explicación generada.

Tabla 9

Rendimiento en anomalía 4

Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1-10)
LIME	100%	Media	Medio	Alta	2/25	9
SHAP	95.3%	Alta	Alto	Media	2/25	7
Árbol de decisión	100%	Baja	Bajo	Muy alta	2/25	10

Fuente: Elaboración de los autores

En este caso, el Árbol de decisión es nuevamente el mejor evaluado (10), al mantener el 100% de precisión junto con una interpretación muy accesible y bajo costo computacional. SHAP muestra una leve caída en precisión (95.3%), lo que, sumado a su alta complejidad, reduce su puntuación a 7. LIME se mantiene como una opción equilibrada con puntuación 9.

5. *Patrón cíclico pregunta–menú en examen*

La *tabla 10* compara el desempeño de LIME, SHAP y Árbol de decisión en la identificación de un patrón de navegación cíclico entre preguntas y menú durante un examen, conducta que podría sugerir búsqueda de respuestas externas. Se evalúan métricas de precisión, complejidad y usabilidad, ofreciendo una visión integral de la idoneidad de cada método para esta anomalía específica.

ISSN: 2745-2891

Tabla 10

Desempeño de anomalía 5

Técnica XAI	Precisión	Complejidad	Tiempo de procesamiento	Facilidad de interpretación	Total detecciones	Puntuación (1-10)
LIME	92.3%	Media	Medio	Alta	2/25	8
SHAP	98.7%	Alta	Alto	Media	2/25	9
Árbol de decisión	91.0%	Baja	Bajo	Muy alta	2/25	8

Fuente: Elaboración de los autores

SHAP obtiene la mayor puntuación (9) en esta anomalía, gracias a su alta precisión (98.7%) y capacidad explicativa media-alta. LIME y el Árbol de decisión registran precisiones ligeramente inferiores (92.3% y 91.0%, respectivamente), lo que se refleja en puntuaciones de 8. Esta anomalía muestra que, en ciertos patrones complejos, métodos como SHAP pueden ofrecer ventajas en detección fina.

Esta *tabla 11* de resumen sintetiza los hallazgos principales del estudio, indicando qué técnica XAI obtuvo la mayor puntuación en cada una de las cinco anomalías analizadas. Proporciona una visión rápida y concluyente sobre la efectividad relativa de LIME, SHAP y Árbol de decisión según el tipo de comportamiento anómalo detectado.

Tabla 11

Resumen comparativo para la elección de la mejor técnica por anomalía

Anomalía	Técnica con mejor puntuación	Puntuación
Entrega de tarea en tiempo insuficiente	Árbol de decisión	10
Publicación masiva en foros	LIME & Árbol de decisión	9
Acceso desde ubicaciones geográficas incompatibles (múltiples IP)	Árbol de decisión	10
Acceso a materiales durante examen activo	Árbol de decisión	10
Patrón cíclico pregunta–menú en examen	SHAP	9

Fuente: Elaboración de los autores

El Árbol de decisión resulta ser la técnica más efectiva en la mayoría de las

anomalías analizadas, especialmente cuando la interpretabilidad y la eficiencia son prioritarias.

SHAP demuestra su fortaleza en anomalías con patrones más complejos, como el ciclo pregunta-menú, donde su precisión fina es determinante. LIME se posiciona como una opción balanceada, útil en escenarios que requieren claridad sin tanta demanda computacional como SHAP.

DISCUSIÓN

La evaluación de un modelo de detección de actividades inusuales no debe limitarse únicamente a su capacidad de clasificación, sino que debe considerar también la utilidad práctica y la claridad de los resultados obtenidos. En el contexto de esta investigación, donde el modelo de árbol de decisiones se plantea como la técnica principal, resulta fundamental emplear métricas que permitan analizar tanto el desempeño del modelo como su nivel de interpretabilidad, por esta razón, se seleccionaron métricas orientadas a evaluar la precisión en la detección de anomalías, la capacidad del modelo para identificar correctamente comportamientos atípicos y la claridad. Estas métricas permiten valorar de manera integral el comportamiento del sistema propuesto y su adecuación al entorno de aulas virtuales universitarias.

CONCLUSIONES

La técnica decision tree de XAI logra el mejor equilibrio entre el desempeño en la detección de anomalías y la calidad de las explicaciones generadas en plataformas educativas. Dicho logro se sustenta en el cumplimiento secuencial de los objetivos específicos delineados desde el inicio del trabajo. Posteriormente, estas métricas fueron aplicadas en un escenario de prueba controlado, lo que permitió medir de forma empírica y objetiva el desempeño de cada técnica tanto en la detección como en la generación de explicaciones útiles y comprensibles. Finalmente, todos los hallazgos fueron organizados, analizados y preparados bajo un formato académico riguroso, sentando las bases para la redacción de un artículo científico destinado a su difusión en una revista especializada de la región, con lo cual se cumple también con el objetivo de divulgación del conocimiento generado.

Los árboles de decisión representan la técnica de inteligencia artificial explicable más adecuada y balanceada para la identificación de

actividades inusuales en entornos de aulas virtuales universitarias. Con una calificación integral de 9.75 sobre 10, este método superó a las demás alternativas evaluadas al conjugar una precisión de detección del 100% con una capacidad explicativa inmediata y accesible, expresada mediante reglas lógicas del tipo SI-ENTONCES. Esta dualidad, exactitud técnica y claridad interpretativa lo hace especialmente valioso en contextos educativos, donde la transparencia y la posibilidad de justificar decisiones ante estudiantes, docentes y autoridades resultan cruciales. Además, su eficiencia computacional y su facilidad de integración en sistemas operativos lo convierten en una solución práctica para escenarios que requieren respuestas en tiempo real. Si bien técnicas como LIME ofrecen explicaciones pedagógicamente más detalladas a nivel individual, y SHAP aporta un rigor formal invaluable para auditorías académicas, el árbol de decisiones emerge como la opción de propósito general más sólida, capaz de cubrir de manera satisfactoria la mayoría de las necesidades detectadas en el ámbito universitario virtual.

En última instancia, la investigación trasciende la mera comparación cuantitativa de algoritmos. Al identificar al árbol de decisiones como la técnica más recomendable para el contexto estudiado, se ofrece a las instituciones de educación superior un marco de referencia práctico y basado en datos que puede guiar la adopción informada de herramientas de inteligencia artificial explicable. Más allá de la detección de irregularidades, este trabajo subraya la importancia de priorizar la interpretabilidad, la confiabilidad y la alineación con los valores educativos en la implementación de sistemas automatizados de monitoreo. De este modo, se contribuye no solo al avance técnico en el campo de la XAI aplicada a la educación, sino también a la construcción de entornos digitales de aprendizaje más transparentes, éticos y pedagógicamente significativos.

REFERENCIAS

Arya, V., Bellamy, R. K., Chen, P. Y., Dhurandhar, A., Hind, M., Hoffman, S. C., ... & Zhang, Y. (2021). AI explainability 360: Impact and future directions. AI Magazine, *42*(2), 8–21.

Baniecki, H., Kretowicz, W., Piatyszek, P., Wisniewski, J., & Biecek, P. (2021). DALEX: Responsible machine learning with interactive explainability and fairness in Python. Journal of Machine Learning Research, *22*(214), 1–7.

Bellas, F., & Kralj, L. (2025). IA explicable en el ámbito educativo. UNESCO – CITIC.

Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., ... & Srivastava, M. (2022). Interpretability of deep learning models: A survey of results. IEEE Access, *10*, 20170–20191.

Chandrasekaran, B., Yadav, D., & Gupta, R. (2022). Explainable AI for anomaly detection in educational data: A comparative study. Computers & Education, *189*, 104–118.

Chávez, D., & Macías, J. (2022). Gestión de seguridad en plataformas virtuales universitarias. Revista Ciencia Digital, *6*(1), 120–134.

Ehsan, U., Wintersberger, P., Liao, Q. V., Mara, M., Streit, M., & Riedl, M. O. (2021). Operationalizing human-centered perspectives in explainable AI. Proceedings of the ACM on Human-Computer Interaction, *5*(CSCW2), 1–36.

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P., & Roth, D. (2021). A comparative study of fairness-enhancing interventions in machine learning. Proceedings of the ACM on Human-Computer Interaction, *5*(CSCW2), 1–35.

Gillies, M., Fiebrink, R., Tanaka, A., Garcia, J., Bevilacqua, F., Heloir, A., ... & Mackay, W. E. (2022). Human-centered machine learning. Interactions, *29*(2), 46–51.

Gummadi, A., Arreche, O., & Abdallan, M. (2025). Una evaluación sistemática de métodos de IA explicables de caja blanca para la detección de anomalías en sistemas IoT. Internet of Things, *30*(1). https://doi.org/10.1016/j.iot.2025.101505

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2021). XAI—Explainable artificial intelligence. Science Robotics, *6*(60), eabg-6261.

Hasan, M., Rahman, M. S., Janicke, H., & Sarker, I. H. (2024). Detecting anomalies in blockchain transactions using machine learning classifiers and explainability analysis. arXiv preprint arXiv:2401.03530.

Kumar, V., & Sharma, A. (2021). Simulated learning environments for cybersecurity education: A systematic review. Education and Information Technologies, *26*(5), 6173–6195.

Liao, Q. V., Gruen, D., & Miller, S. (2022). Questioning the AI: Informing design practices for explainable AI user experiences. Proceedings of the ACM on Human-Computer Interaction, *6*(CSCW1), 1–27.

Lundberg, S. M., & Lee, S. I. (2020). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, *33*, 4765–4774.

Maricar, A., Anoop, A., Samuel, B. E., & Alsinjlawi, K. H. (2024). An improved explainable artificial intelligence for intrusion detection system. International Journal of Intelligent Systems and Applications in Engineering, *12*(3), 108–115.

Ministerio de Telecomunicaciones y de la Sociedad de la Información (MINTEL). (2022). Plan Nacional de Transformación Digital 2022–2025. Gobierno de Ecuador.

Misra, S., Li, H., & He, J. (2021). Noninvasive fracture characterization based on the classification of sonic wave travel times. Machine Learning for Science and Engineering, *2*(1), 1–23.

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable machine learning–A brief history, state-of-the-art and challenges. arXiv preprint arXiv:2010.09337.

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, *115*, 619–640.

Pandey, R., & Tripathi, A. K. (2022). Design and development of a simulated virtual learning environment for cybersecurity training. International Journal of

Emerging Technologies in Learning, *17*(8), 234–250.

Reinoso, P., & Morales, S. (2023). Evaluación de estrategias de seguridad informática en entornos educativos digitales. Repositorio Institucional Universidad de Guayaquil.

Reyes, M., & Toala, C. (2023). Gestión de accesos y control de datos en plataformas tecnológicas educativas. Repositorio Institucional de la Universidad de Guayaquil.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2020). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

Rudin, C. (2020). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, *2*(5), 206–215.

Singh, A., Thakur, N., & Sharma, A. (2022). A review of supervised machine learning algorithms for heart disease prediction. International Journal of Computer Applications, *183*(41), 31–37.

Suresh, H., Gomez, S. R., Nam, K. K., & Satyanarayan, A. (2021). Beyond expertise and roles: A framework to characterize the stakeholders of interpretable machine learning and their needs. Proceedings of the ACM on Human-Computer Interaction, *5*(CSCW1), 1–32.

UNESCO. (2024). Directrices para el uso ético y explicable de la inteligencia artificial en educación. Oficina de Publicaciones de la UNESCO.

Valencia, L., Aguirre, C., & Cedeño, M. (2023). Gestión de la seguridad de la información en aulas virtuales universitarias. Revista Ciencia Digital, *7*(2), 45–61.

Verma, S., Dickerson, J., & Hines, K. (2023). Counterfactual explanations for machine learning: A review. arXiv preprint arXiv:2301.03025.

Zhang, Y., Tino, P., Leonardis, A., & Tang, K. (2021). A survey on neural network interpretability. IEEE Transactions on Emerging Topics

in Computational Intelligence, *5*(5), 726–742.

Zhao, X., Wu, Y., Lee, D. L., & Cui, W. (2023). A survey on evaluation methods for explainable AI. ACM Computing Surveys, *56*(3), 1–40.