Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), marzo-abril 2026,
Volumen 10, Número 2.

https://doi.org/10.37811/cl_rcm.v10i2

MODELOS PREDICTIVOS DE SEGURIDAD BASADOS EN INTELIGENCIA ARTIFICIAL

PREDICTIVE SECURITY MODELS BASED ON ARTIFICIAL INTELLIGENCE

María Teodolinda Ortega Ovalle
Universidad de Panamá - Panamá

DOI: https://doi.org/10.37811/cl_rcm.v10i2.23654

Modelos predictivos de seguridad basados en inteligencia artificial

María Teodolinda Ortega Ovalle[1]
maria.ortegao@up.ac.pa
https://orcid.org/0009-0000-3629-9751
Universidad de Panamá
Panamá

RESUMEN

El presente trabajo tiene como objetivo analizar el papel de los modelos predictivos de seguridad basados en inteligencia artificial como herramientas estratégicas para anticipar amenazas y reducir riesgos en entornos digitales. La metodología implementada se fundamenta en una revisión documental de literatura científica reciente, complementada con el estudio de casos prácticos en ciberseguridad empresarial, seguridad pública y protección de infraestructuras críticas. Se emplearon técnicas de análisis comparativo para identificar patrones comunes en la aplicación de algoritmos supervisados, no supervisados y de aprendizaje profundo, así como sus ventajas y limitaciones en la detección de anomalías y la prevención de ataques. Los principales hallazgos evidencian que la inteligencia artificial permite transformar la seguridad de un enfoque reactivo hacia uno proactivo, mejorando la capacidad de respuesta y reduciendo la exposición a vulnerabilidades. Sin embargo, se identifican retos asociados a la calidad de los datos, la interpretabilidad de los modelos y la gestión de falsos positivos, lo que plantea la necesidad de un equilibrio entre innovación tecnológica y gobernanza ética.

***Palabras clave*:** Inteligencia artificial; modelos predictivos; ciberseguridad; detección de anomalías; seguridad proactiva

[1] Autor Principal
Correspondencia: maria.ortegao@up.ac.pa

Predictive security models based on artificial intelligence

ABSTRACT

This study aims to analyze the role of predictive security models based on artificial intelligence as strategic tools to anticipate threats and reduce risks in digital environments. The methodology implemented relies on a documentary review of recent scientific literature, complemented by case studies in corporate cybersecurity, public safety, and critical infrastructure protection. Comparative analysis techniques were applied to identify common patterns in the use of supervised, unsupervised, and deep learning algorithms, as well as their advantages and limitations in anomaly detection and attack prevention. The main findings show that artificial intelligence enables a shift from reactive to proactive security, enhancing response capacity and reducing exposure to vulnerabilities. However, challenges related to data quality, model interpretability, and false positive management were identified, highlighting the need for a balance between technological innovation and ethical governance.

***Keywords*:** Artificial intelligence, predictive models, cybersecurity, anomaly detection, proactive security

Artículo recibido 20 marzo 2026
Aceptado para publicación: 15 abril 2026

INTRODUCCIÓN

La seguridad digital constituye uno de los retos más significativos de la sociedad contemporánea. En un mundo cada vez más interconectado, donde los procesos sociales, económicos y políticos dependen de sistemas digitales, la protección de la información y de las infraestructuras críticas se ha convertido en una prioridad estratégica. El tema que se aborda en este artículo son los modelos predictivos de seguridad basados en inteligencia artificial (IA), entendidos como herramientas que permiten anticipar amenazas y reducir riesgos mediante el análisis de datos históricos y en tiempo real.

El problema de investigación que se plantea surge de la insuficiencia de los enfoques tradicionales de seguridad, que suelen ser reactivos y dependen de la detección posterior al ataque. Estos métodos, aunque útiles en determinados contextos, resultan limitados frente a la sofisticación de los ciberataques actuales, que evolucionan con rapidez y aprovechan vulnerabilidades antes de que puedan ser corregidas. El vacío en el conocimiento se traduce en la necesidad de explorar cómo la IA puede transformar la seguridad hacia un enfoque proactivo, capaz de prever incidentes antes de que ocurran y, en consecuencia, reducir la exposición a riesgos.

La relevancia de este estudio se justifica en la magnitud de las amenazas digitales que enfrentan tanto organizaciones privadas como instituciones públicas. Los ataques de ransomware, phishing, denegación de servicio distribuida (DDoS) y explotación de vulnerabilidades en software representan riesgos que afectan la continuidad de los servicios, la integridad de los datos y la confianza de los usuarios. Abordar este tema resulta esencial para garantizar la resiliencia organizacional y social, así como para fortalecer la gobernanza digital en un entorno global cada vez más dependiente de la tecnología.

El marco teórico de este trabajo se sustenta en los postulados del aprendizaje automático (machine learning) y del aprendizaje profundo (deep learning), que permiten construir modelos capaces de identificar patrones ocultos y anomalías en grandes volúmenes de datos. Autores como Bishop (2006) han señalado que los algoritmos supervisados, tales como la regresión logística y los árboles de decisión, son útiles para clasificar eventos de seguridad conocidos. Por su parte, Goodfellow, Bengio y Courville (2016) destacan que las redes neuronales profundas ofrecen ventajas en la detección de comportamientos complejos y en la predicción de ataques emergentes. Las categorías de análisis utilizadas en este trabajo

incluyen la detección de anomalías, la predicción de vulnerabilidades y la gestión de riesgos, todas ellas fundamentales para comprender el potencial de los modelos predictivos en seguridad.

En cuanto a los antecedentes investigativos, diversos estudios han demostrado la eficacia de los modelos predictivos en la detección temprana de ataques. Shahid et al. (2020) evidencian que el uso de algoritmos de clustering no supervisado permite identificar patrones de tráfico anómalo asociados a intentos de intrusión. Alazab et al. (2021) señalan que las redes neuronales recurrentes (RNN) aplicadas a series temporales de tráfico de red mejoran la capacidad de predicción de ataques DDoS. Sin embargo, los estudios también advierten sobre limitaciones relacionadas con la calidad de los datos, la interpretabilidad de los modelos y la generación de falsos positivos, lo que plantea la necesidad de un abordaje integral que combine innovación tecnológica con marcos éticos y regulatorios.

El contexto de la investigación se sitúa en un escenario global caracterizado por el aumento exponencial de los ciberataques y la creciente dependencia de tecnologías digitales en sectores estratégicos como la banca, la salud, la energía y el transporte. En América Latina, por ejemplo, el incremento de ataques de ransomware ha puesto de manifiesto la vulnerabilidad de las instituciones públicas y privadas, mientras que en Europa y Estados Unidos se han registrado incidentes que afectan directamente a infraestructuras críticas. Este entorno evidencia la urgencia de implementar modelos predictivos de seguridad que permitan anticipar amenazas y fortalecer la resiliencia organizacional.

La presente investigación se propone analizar los fundamentos teóricos y metodológicos de los modelos predictivos de seguridad basados en inteligencia artificial, así como sus principales aplicaciones y limitaciones. Para ello, se ha implementado una estrategia metodológica de revisión documental de literatura científica reciente, complementada con el estudio de casos prácticos en ciberseguridad empresarial, seguridad pública y protección de infraestructuras críticas. El análisis comparativo de estas experiencias permite identificar patrones comunes en la implementación de modelos predictivos, así como las condiciones necesarias para su éxito.

Los hallazgos preliminares sugieren que la integración de modelos predictivos en los sistemas de seguridad contribuye a mejorar la resiliencia organizacional, al reducir la exposición a vulnerabilidades y fortalecer la capacidad de respuesta frente a ataques emergentes. No obstante, se reconoce que la eficacia de estos modelos depende en gran medida de la calidad de los datos utilizados, de la capacidad

de los equipos humanos para interpretar los resultados y de la existencia de marcos éticos y regulatorios que orienten su aplicación. En este sentido, la investigación busca aportar una reflexión crítica sobre el papel de la inteligencia artificial en la seguridad contemporánea, destacando tanto sus beneficios como los desafíos que plantea.

Finalmente, el objetivo general de este estudio es analizar los fundamentos, aplicaciones y limitaciones de los modelos predictivos de seguridad basados en inteligencia artificial, con el propósito de demostrar su potencial para transformar la gestión de riesgos en entornos digitales. Los objetivos específicos incluyen: (a) describir los componentes técnicos de los modelos predictivos, (b) identificar sus principales beneficios y desafíos, y (c) examinar casos prácticos de aplicación en distintos sectores.

METODOLOGÍA

El presente estudio se desarrolló bajo un enfoque mixto, combinando elementos cuantitativos y cualitativos con el propósito de obtener una visión integral sobre la aplicación de modelos predictivos de seguridad basados en inteligencia artificial. El componente cuantitativo permitió sistematizar y comparar datos provenientes de investigaciones previas y casos documentados, mientras que el componente cualitativo se centró en el análisis interpretativo de literatura científica y en la identificación de categorías conceptuales relevantes para el tema.

En cuanto al tipo de investigación, se adoptó un diseño exploratorio-descriptivo, dado que el objetivo principal fue analizar los fundamentos teóricos y metodológicos de los modelos predictivos, describir sus aplicaciones en distintos sectores y explorar sus limitaciones. Este tipo de investigación resulta pertinente en campos emergentes como la ciberseguridad, donde aún existen vacíos de conocimiento y se requiere sistematizar experiencias previas para orientar futuros desarrollos.

El diseño metodológico fue de carácter observacional y transversal, ya que se trabajó con fuentes secundarias de información y se realizó un análisis en un momento determinado, sin manipulación de variables ni intervención experimental. La revisión documental incluyó artículos científicos indexados en bases de datos como Scopus, IEEE Xplore y Web of Science, publicados entre 2018 y 2024, lo que permitió garantizar la actualidad y relevancia de los hallazgos.

La población de estudio estuvo conformada por investigaciones académicas y reportes técnicos relacionados con la aplicación de inteligencia artificial en seguridad digital. Se establecieron criterios

de inclusión que consideraron estudios publicados en revistas científicas de acceso abierto o indexadas, documentos con enfoque explícito en modelos predictivos y trabajos que presentaran resultados empíricos o aplicaciones prácticas. Los criterios de exclusión descartaron publicaciones sin rigor metodológico, artículos de divulgación sin sustento científico y documentos anteriores al año 2018, por considerarse menos representativos del estado actual del campo.

Las técnicas de recolección de datos se basaron en la revisión documental sistemática, siguiendo protocolos de búsqueda y selección de información. Se utilizaron palabras clave como inteligencia artificial, modelos predictivos, ciberseguridad, detección de anomalías y seguridad proactiva, tanto en español como en inglés, para ampliar el espectro de resultados. El instrumento principal fue una matriz de análisis comparativo, en la que se registraron las características metodológicas, los algoritmos utilizados, los resultados obtenidos y las limitaciones señaladas en cada estudio.

En el componente cualitativo, se aplicó un análisis de contenido temático, que permitió identificar categorías recurrentes en la literatura, tales como la eficacia de los modelos, los retos de interpretabilidad, la dependencia de datos de calidad y las consideraciones éticas. Este procedimiento facilitó la construcción de un marco conceptual sólido y la integración de hallazgos dispersos en una visión coherente.

Respecto a las consideraciones éticas, se garantizó el respeto a la propiedad intelectual mediante la citación adecuada de las fuentes consultadas, siguiendo las normas del Manual de Estilo de la American Psychological Association (APA, séptima edición). Asimismo, se evitó cualquier sesgo en la selección de estudios, procurando incluir investigaciones de diferentes regiones y enfoques metodológicos para asegurar la diversidad de perspectivas.

Finalmente, se reconocen algunas limitaciones del estudio. Al tratarse de una investigación documental, los resultados dependen de la disponibilidad y calidad de las publicaciones existentes, lo que puede restringir la generalización de los hallazgos. Además, la ausencia de un componente experimental impide validar directamente la eficacia de los modelos predictivos en escenarios reales, aunque se compensa con el análisis comparativo de casos documentados en la literatura.

Marco Teórico

El marco teórico de los modelos predictivos de seguridad basados en inteligencia artificial se fundamenta en el aprendizaje automático y profundo, la teoría de la detección de anomalías y la gestión de riesgos en ciberseguridad. Estos enfoques permiten pasar de una seguridad reactiva a una proactiva, anticipando amenazas antes de que se materialicen. La inteligencia artificial se define como la capacidad de sistemas computacionales para realizar tareas que requieren inteligencia humana, como el reconocimiento de patrones, la toma de decisiones y el aprendizaje adaptativo. Dentro de ella, el aprendizaje automático constituye un conjunto de algoritmos que permiten a los sistemas aprender de datos históricos y mejorar su desempeño sin ser programados explícitamente, mientras que el aprendizaje profundo utiliza redes neuronales para identificar patrones complejos en grandes volúmenes de datos, aplicables en la detección de anomalías y predicción de ataques.

La teoría de la detección de anomalías se basa en la premisa de que los ataques generan comportamientos distintos al tráfico normal. Algoritmos como Support Vector Machines y Random Forest se utilizan para clasificar eventos sospechosos, mientras que técnicas de clustering como K-means permiten identificar patrones ocultos en los datos. En paralelo, la gestión de riesgos aplicada a la inteligencia artificial en seguridad se apoya en marcos como MITRE ATLAS, OWASP y el NIST AI Risk Management, que enfatizan la necesidad de transparencia, interpretabilidad y mitigación de sesgos en los modelos predictivos. Estos marcos teóricos aportan lineamientos para evaluar riesgos asociados al uso de IA en ciberseguridad y garantizan que las soluciones tecnológicas se implementen de manera ética y responsable.

Los antecedentes investigativos muestran que la aplicación de modelos predictivos en seguridad digital ha tenido resultados significativos. Shahid et al. (2020) evidenciaron que el uso de clustering no supervisado permite identificar patrones de tráfico anómalo asociados a intentos de intrusión. Alazab et al. (2021) demostraron que las redes neuronales recurrentes aplicadas a series temporales de tráfico de red incrementan la capacidad predictiva frente a ataques de denegación de servicio distribuida. Castro Peña (2022) señala que la inteligencia artificial aplicada a la seguridad enfrenta retos de interpretabilidad y sobre entrenamiento, pero ofrece un potencial considerable en la prevención de delitos digitales. Estos estudios confirman que los modelos predictivos pueden mejorar la capacidad de detección temprana de

amenazas, aunque también advierten sobre limitaciones relacionadas con la calidad de los datos y la generación de falsos positivos.

Las variables técnicas que se consideran en este marco incluyen el tipo de algoritmo utilizado, el volumen y calidad de los datos y la tasa de falsos positivos y negativos. Las categorías de análisis abarcan la detección de anomalías, la predicción de vulnerabilidades, la gestión de riesgos y la interpretabilidad de los modelos. Asimismo, los factores contextuales como la protección de infraestructuras críticas, la seguridad empresarial y la seguridad pública resultan esenciales para comprender la pertinencia de los modelos predictivos en distintos escenarios.

En síntesis, el marco teórico de los modelos predictivos de seguridad basados en inteligencia artificial se apoya en tres pilares fundamentales: los fundamentos del aprendizaje automático y profundo para la detección y predicción de amenazas, las teorías de gestión de riesgos y resiliencia que orientan la aplicación segura y ética de la IA, y los antecedentes investigativos recientes que evidencian tanto el potencial como las limitaciones de estos modelos. En conjunto, este marco proporciona la base conceptual para comprender cómo la inteligencia artificial puede transformar la seguridad digital, pasando de un enfoque reactivo a uno predictivo y proactivo, y ofreciendo nuevas posibilidades para enfrentar los desafíos de la ciberseguridad contemporánea.

RESULTADOS Y DISCUSIÓN

Los resultados de la investigación documental y comparativa sobre modelos predictivos de seguridad basados en inteligencia artificial evidencian que estos sistemas constituyen una herramienta eficaz para anticipar amenazas y reducir riesgos en entornos digitales. La revisión de literatura permitió identificar patrones comunes en la aplicación de algoritmos supervisados, no supervisados y de aprendizaje profundo, así como sus ventajas y limitaciones en distintos contextos.

En primer lugar, se observó que los modelos supervisados, como la regresión logística y los árboles de decisión, son útiles para clasificar eventos de seguridad conocidos, aunque presentan limitaciones frente a ataques novedosos. Los modelos no supervisados, como el clustering, mostraron mayor capacidad para detectar anomalías en tráfico de red, mientras que las redes neuronales profundas ofrecieron resultados superiores en la predicción de ataques complejos, especialmente en escenarios de gran volumen de datos.

En segundo lugar, los hallazgos sugieren que la eficacia de los modelos predictivos depende en gran medida de la calidad y representatividad de los datos utilizados. Los estudios revisados coinciden en que la presencia de datos incompletos o sesgados puede afectar la precisión de las predicciones y aumentar la tasa de falsos positivos. Asimismo, se identificó que la interpretabilidad de los modelos constituye un reto significativo, ya que muchos algoritmos avanzados funcionan como "cajas negras" difíciles de explicar para los equipos de seguridad.

En tercer lugar, se constató que la integración de modelos predictivos en sistemas de seguridad empresarial y en infraestructuras críticas contribuye a mejorar la resiliencia organizacional. Sin embargo, también se reconocen desafíos relacionados con la gobernanza ética, la protección de la privacidad y la necesidad de marcos regulatorios que orienten su aplicación.

Para ilustrar los hallazgos, se presentan a continuación cuatro tablas que sintetizan los principales resultados:

Tabla 1. Tipos de algoritmos aplicados en modelos predictivos de seguridad

Tipo de algoritmo	Ejemplos	Aplicación principal	Limitaciones principales
Supervisados	Regresión logística, SVM	Clasificación de ataques conocidos	Baja eficacia frente a ataques nuevos
No supervisados	K-means, DBSCAN	Detección de anomalías en tráfico	Sensibilidad a parámetros de clustering
Aprendizaje profundo	RNN, CNN	Predicción de ataques complejos	Interpretabilidad limitada, alto costo

Tabla 2. Beneficios de los modelos predictivos de seguridad

Beneficio	Descripción
Mitigación proactiva de riesgos	Anticipa vulnerabilidades antes de ser explotadas
Aprendizaje continuo	Los modelos mejoran con cada nuevo dato analizado
Reducción del tiempo de respuesta	Permite actuar antes de que el ataque se materialice
Optimización de recursos	Disminuye costos asociados a incidentes de seguridad

Tabla 3. Retos y limitaciones identificados

Reto/Limitación	Impacto en la eficacia del modelo
Calidad de los datos	Datos incompletos o sesgados reducen precisión
Interpretabilidad	Dificultad para explicar decisiones de modelos complejos
Falsos positivos/negativos	Generan desconfianza en el sistema
Gobernanza ética	Riesgos de sesgo algorítmico y afectación de la privacidad

Tabla 4. Aplicaciones prácticas de los modelos predictivos

Sector	Ejemplo de aplicación	Resultados observados
Ciberseguridad empresarial	Integración en sistemas SIEM e IDS/IPS	Mejora en detección temprana de ataques
Seguridad pública	Análisis predictivo de delitos	Prevención de incidentes en zonas críticas
Infraestructuras críticas	Protección de sistemas de energía y transporte	Reducción de vulnerabilidades

En conclusión, los resultados muestran que los modelos predictivos de seguridad basados en inteligencia artificial ofrecen un potencial significativo para transformar la gestión de riesgos en entornos digitales. No obstante, su implementación requiere superar retos técnicos y éticos, así como garantizar la disponibilidad de datos de calidad y la existencia de marcos regulatorios adecuados. La discusión evidencia que el futuro de la seguridad digital dependerá de la capacidad de integrar estos modelos en estrategias organizacionales más amplias, orientadas a la resiliencia y a la protección integral de los sistemas.

CONCLUSIONES

El presente estudio permitió analizar de manera integral los fundamentos, aplicaciones y limitaciones de los modelos predictivos de seguridad basados en inteligencia artificial. Los hallazgos evidencian que estos modelos constituyen una herramienta estratégica para transformar la gestión de riesgos en entornos

digitales, al pasar de un enfoque reactivo hacia uno proactivo. La revisión documental y el análisis comparativo de casos demostraron que los algoritmos supervisados, no supervisados y de aprendizaje profundo ofrecen capacidades diferenciadas para la detección de anomalías y la predicción de ataques, siendo los modelos de aprendizaje profundo los que presentan mayor eficacia en escenarios de gran volumen de datos.

Asimismo, se constató que la calidad y representatividad de los datos son factores determinantes para la precisión de las predicciones, mientras que la interpretabilidad de los modelos continúa siendo un reto que limita su adopción en algunos contextos. La integración de estos sistemas en la seguridad empresarial, la protección de infraestructuras críticas y la seguridad pública ha mostrado resultados positivos, aunque también plantea desafíos relacionados con la gobernanza ética, la privacidad y la necesidad de marcos regulatorios que orienten su aplicación responsable.

En términos prácticos, los modelos predictivos de seguridad basados en inteligencia artificial contribuyen a mejorar la resiliencia organizacional, optimizar recursos y reducir el tiempo de respuesta frente a amenazas emergentes. Sin embargo, su implementación requiere un equilibrio entre innovación tecnológica y principios éticos, así como la capacitación de los equipos humanos para interpretar y gestionar los resultados de manera adecuada.

Finalmente, se concluye que la inteligencia artificial aplicada a la seguridad digital no solo representa una oportunidad para fortalecer la defensa frente a ataques cada vez más sofisticados, sino también un campo de investigación en constante evolución que demanda estudios adicionales. Futuras investigaciones deberán profundizar en la creación de modelos más interpretables, en la reducción de falsos positivos y en la integración de marcos regulatorios internacionales que garanticen su uso seguro y equitativo. De esta manera, los modelos predictivos de seguridad basados en inteligencia artificial podrán consolidarse como un pilar fundamental en la protección de los sistemas digitales y en la construcción de sociedades más seguras y resilientes.

REFERENCIAS BIBLIOGRÁFICAS

Ahmad, Z., Khan, A. S., Shiang, C. W., Abdullah, J., & Ahmad, F. (2021). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies, 32*(1), e4150. https://doi.org/10.1002/ett.4150

Alashhab, A. A., Zahid, M. S. M., Muneer, A., & Abdullahi, M. (2024, enero). Detección de ataques DDoS a baja tasa usando Deep Learning para redes IoT habilitadas con SDN [Preprint]. *TechRxiv*. https://doi.org/10.36227/techrxiv.170472893.31491821/v1

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. ISBN: 978-0387310732

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

Castro Peña, J. L. (2023). Inteligencia artificial aplicada a la seguridad. *Logos Guardia Civil: Revista Científica del Centro Universitario de la Guardia Civil, (1)*, 35–60. https://revistacugc.es/article/view/5911

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. ISBN: 9780262035613

Shaukat, K., Luo, S., Varadharajan, V., & Hameed, I. A. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies, 13*(10), 2509. https://doi.org/10.3390/en13102509

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316. https://doi.org/10.1109/SP.2010.26

Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized smartphone usage in a context-aware environment. *Journal of Big Data, 6*(1), 57. https://doi.org/10.1186/s40537-019-0219-y

Tiwari, S., & Kumar, A. (2024). Novel bit-level adaptive and asymmetric data compression technique [Preprint]. *TechRxiv*. School of Computer Science, University of Petroleum and Energy Studies (UPES), Dehradun. https://doi.org/10.36227/techrxiv.170472893.31491821/v1