ASCE MAGAZINE ISSN: 3073-1178

Revista ASCE Magazine, Periodicidad: Trimestral Julio - Septiembre, Volumen: 4, Número: 3, Año: 2025 páginas 333-352

Doi: https://doi.org/10.70577/ASCE/333.352/2025

Recibido: 2025-05-09

Aceptado:2025-06-09

Publicado:2025-07-11

¿Puede el Aprendizaje Automático Predecir Brechas de Ciberseguridad en los Sistemas de Información Empresariales? Un análisis de aprendizaje supervisado

Can Machine Learning Predict Cybersecurity Breaches in Enterprise Information Systems? A Supervised Learning Analysis.

Autor:

Sonia Patricia Córdovez Machado
https://orcid.org/0000-0002-2393-7918
sonia.cordovez@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo (ESPOCH)
Riobamba-Ecuador

John Javier Cruz Garzón
https://orcid.org/0009-0006-7550-9962
jjcruz9@espe.edu.ec
Universidad de las Fuerzas Armadas (ESPE)
Santo Domingo-Ecuador

Cristian Luis Inca Balseca
https://orcid.org/0000-0002-4795-8297
cristianl.inca@espoch.edu.ec
Escuela Superior Politécnica de Chimborazo (ESPOCH)
Riobamba-Ecuador

Cómo citar

Córdovez Machado, S. P., Cruz Garzón, J. J., & Inca Balseca, C. L. (2025). ¿Puede el Aprendizaje Automático Predecir Brechas de Ciberseguridad en los Sistemas de Información Empresariales? Un análisis de aprendizaje supervisado . *ASCE*, *4*(3), 333–352.

Resumen

Este artículo investiga la capacidad del aprendizaje automático supervisado para predecir brechas de ciberseguridad en entornos empresariales. Mediante un análisis comparativo de dos modelos, la Máquina de Vectores de Soporte (SVM) y el Bosque Aleatorio (Random Forest), el estudio evalúa su eficacia en un conjunto de datos simulado que integra variables técnicas, contextuales y de comportamiento humano. Ambos modelos alcanzan una notable precisión global del 88%, aunque con fortalezas distintas: el SVM destaca por su alta sensibilidad para detectar brechas reales, mientras que el Bosque Aleatorio demuestra una consistencia y fiabilidad superiores. El hallazgo más significativo proviene del análisis de importancia de variables del Bosque Aleatorio, que revela que el factor humano —representado por la tasa de clics en phishing— es el predictor más influyente, acaparando el 42% del poder predictivo. Este supera considerablemente a los factores técnicos como la gravedad o el número de vulnerabilidades. El estudio concluye que, si bien el aprendizaje automático es una herramienta potente para la predicción, su mayor valor reside en identificar los verdaderos focos de riesgo. Por ello, se recomienda que las estrategias de ciberseguridad se reorienten para priorizar la mitigación del riesgo humano, reconociéndolo no solo como una vulnerabilidad, sino como el principal indicador predictivo de un incidente.

Palabras clave: Ciberseguridad; Aprendizaje Automático; Análisis Predictivo; Factor Humano; Bosque Aleatorio (Random Forest).

Abstract

This article investigates the ability of supervised machine learning to predict cybersecurity breaches in enterprise environments. Through a comparative analysis of two models, the Support Vector Machine (SVM) and the Random Forest, the study evaluates their effectiveness on a simulated dataset that integrates technical, contextual, and human behavioral variables. Both models achieve a remarkable overall accuracy of 88%, albeit with different strengths: the SVM stands out for its high sensitivity in detecting real breaches, while the Random Forest demonstrates superior consistency and reliability. The most significant finding comes from the variable importance analysis of the Random Forest, which reveals that the human factor—represented by the phishing click-through rate—is the most influential predictor, accounting for 42% of the predictive power. This considerably outperforms technical factors such as severity or number of vulnerabilities. The study concludes that, while machine learning is a powerful tool for prediction, its greatest value lies in identifying true risk hotspots. Therefore, it is recommended that cybersecurity strategies be reoriented to prioritize human risk mitigation, recognizing it not only as a vulnerability, but as the main predictive indicator of an incident.

Keywords: Cybersecurity; Machine Learning; Predictive Analytics; Human Factor; Random Forest.

Introducción

¿Puede el Aprendizaje Automático Predecir las Brechas de Ciberseguridad en los Sistemas de Información Empresariales? Un análisis de aprendizaje supervisado explora la intersección del aprendizaje automático (ML) y la ciberseguridad, centrándose en el potencial de las metodologías de aprendizaje supervisado para predecir y mitigar las brechas de ciberseguridad en los sistemas de información empresarial.

A medida que las amenazas cibernéticas continúan evolucionando en complejidad y frecuencia, la urgencia para que las organizaciones adopten medidas predictivas robustas nunca ha sido mayor. Este análisis destaca la importancia del ML en la mejora de la detección y respuesta a amenazas, convirtiéndolo en un área notable de estudio tanto en los ámbitos académicos como prácticos de la ciberseguridad. El artículo profundiza en los fundamentos del aprendizaje automático, detallando sus diversos tipos, incluyendo el aprendizaje supervisado, no supervisado y por refuerzo.

En particular, se enfatiza el aprendizaje supervisado por su papel en el entrenamiento de algoritmos con conjuntos de datos etiquetados, permitiéndoles discernir entre actividades benignas y maliciosas. Esta capacidad es crucial para las empresas que buscan abordar proactivamente posibles brechas de seguridad mediante la predicción de vectores de ataque basados en patrones de datos históricos.

El análisis también revisa las metodologías involucradas en el aprendizaje supervisado, como la preprocesamiento de datos, la extracción de características y el entrenamiento de modelos, todas las cuales contribuyen a la efectividad general de la analítica predictiva en ciberseguridad.

Las controversias prominentes en torno a la aplicación del aprendizaje automático en la ciberseguridad incluyen desafíos relacionados con la calidad de los datos, el rendimiento del modelo y el potencial de ataques adversariales. Los críticos argumentan que los sesgos en los conjuntos de datos de entrenamiento pueden llevar a predicciones inexactas, mientras que los problemas de sobreajuste y subajuste pueden comprometer la fiabilidad del modelo.

Además, la susceptibilidad de los sistemas de ML a la manipulación por actores maliciosos plantea preocupaciones sobre la integridad de las medidas de seguridad automatizadas. Estos desafíos destacan la necesidad de mejoras continuas y técnicas de validación robustas para mejorar la efectividad del aprendizaje automático en la lucha contra las amenazas de ciberseguridad. En general, el análisis subraya el potencial transformador del aprendizaje automático en la predicción y mitigación de brechas de ciberseguridad, al tiempo que reconoce las limitaciones inherentes y las consideraciones éticas asociadas con su implementación.

Al aprovechar técnicas avanzadas de aprendizaje automático, las organizaciones pueden fortalecer significativamente sus defensas, agilizar los esfuerzos de respuesta a incidentes y navegar por el paisaje cada vez más complejo de las amenazas cibernéticas.

Fundamentos del Aprendizaje Automático

El aprendizaje automático (ML) es un subconjunto de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos que permiten a las computadoras aprender de los datos y hacer predicciones o tomar decisiones sin programación explícita. Este enfoque permite a los sistemas reconocer patrones en los datos y mejorar su rendimiento con el tiempo a medida que procesan más información (May, 2024).

Tipos de Aprendizaje Automático

Hay tres tipos principales de aprendizaje automático, cada uno con metodologías y aplicaciones distintas:

Aprendizaje Supervisado

En el aprendizaje supervisado, los algoritmos se entrenan con datos etiquetados, donde los datos de entrada están emparejados con la salida correspondiente. Este entrenamiento permite al modelo predecir resultados para datos nuevos y no vistos. Las aplicaciones comunes en ciberseguridad incluyen clasificar muestras como benignas o maliciosas, lo que permite la detección de amenazas. Algoritmos como los árboles de decisión, las máquinas de soporte vectorial (SVM) y las redes neuronales se utilizan frecuentemente en este contexto (May, 2024).

Aprendizaje No Supervisado

El aprendizaje no supervisado implica entrenar modelos con datos no etiquetados, permitiéndoles identificar patrones y estructuras dentro de los datos de manera autónoma. Esta técnica es particularmente útil para la detección de anomalías y el descubrimiento de nuevos patrones de ataque en ciberseguridad, ayudando a identificar amenazas y comportamientos previamente desconocidos (May, 2024).

Aprendizaje por Refuerzo

El aprendizaje por refuerzo es un tipo de ML donde los modelos aprenden interactuando con su entorno, recibiendo recompensas por acciones correctas y penalizaciones por acciones incorrectas. Este enfoque imita de cerca el aprendizaje humano y es beneficioso para tareas complejas como los sistemas autónomos de detección de intrusiones y la gestión de ataques distribuidos de denegación de servicio (DDoS) (May, 2024).

Importancia del Aprendizaje Automático en la Ciberseguridad

El aprendizaje automático desempeña un papel crucial en las estrategias modernas de ciberseguridad, principalmente debido a su capacidad para abordar la creciente sofisticación de las amenazas cibernéticas. Al aprovechar el aprendizaje automático, las organizaciones pueden automatizar la detección de anomalías, mejorar los tiempos de respuesta ante amenazas y mejorar los protocolos de seguridad en general (Wang, 2023).

Además, las técnicas de ML, como el análisis predictivo y el procesamiento del lenguaje natural, se han integrado en los marcos de ciberseguridad existentes, fortaleciendo aún más las defensas contra actividades maliciosas.

Incidentes de Ciberseguridad

Las brechas de ciberseguridad se refieren a incidentes en los que se produce un acceso no autorizado a una infraestructura informática, lo que lleva a la interrupción, robo o manipulación de datos y sistemas sensibles. Estas brechas pueden tener consecuencias devastadoras para individuos, empresas y gobiernos por igual, a menudo resultando en pérdidas financieras, daños a la reputación y repercusiones legales. A medida que los ciberataques aumentan en frecuencia y

sofisticación, las organizaciones deben adoptar medidas de seguridad robustas para proteger sus activos e información.

Tipos de Brechas de Ciberseguridad

Las brechas de ciberseguridad pueden manifestarse de diversas formas, cada una con metodologías e impactos distintos:

Ataques de Malware

El malware abarca varios tipos de software malicioso diseñados para infiltrarse en los sistemas y realizar acciones dañinas. Esto incluye el ransomware, que cifra los datos para extorsionar un pago, y el spyware, que monitorea secretamente las actividades del usuario. El malware típicamente entra en un sistema a través de enlaces o archivos adjuntos engañosos (Jada, 2024).

Ataques de Phishing

El phishing implica que los ciberdelincuentes se hagan pasar por entidades de confianza para engañar a las víctimas y que estas revelen información confidencial, como contraseñas o números de tarjetas de crédito. Esto se lleva a cabo típicamente a través de correos electrónicos o mensajes fraudulentos que parecen legítimos.

Ataques de Denegación de Servicio (DoS)

Los ataques DoS tienen como objetivo inutilizar un sistema al abrumarlo con tráfico. Una variante distribuida, conocida como Denegación de Servicio Distribuida (DDoS), emplea múltiples dispositivos comprometidos para lanzar un ataque coordinado, complicando aún más los esfuerzos de mitigación.

Consecuencias de las Brechas de Ciberseguridad

Las ramificaciones de las brechas de ciberseguridad son extensas:

Impacto Económico

El impacto financiero de una violación puede incluir el robo de información sensible, interrupciones en las operaciones comerciales y los costos asociados con la restauración del

sistema. Las empresas también pueden incurrir en sanciones y honorarios legales debido a la falta de cumplimiento normativo relacionado con las leyes de protección de datos (Jada, 2024).

Daño Reputacional

Las brechas pueden erosionar significativamente la confianza del cliente, lo que lleva a una disminución de las ventas y un daño a largo plazo a la reputación de la marca. A medida que los clientes se vuelven cada vez más conscientes de los problemas de ciberseguridad, su lealtad puede disminuir tras un incidente.

Consecuencias Legales

Las organizaciones pueden enfrentar consecuencias legales si no protegen adecuadamente los datos personales. Las medidas de seguridad inadecuadas pueden llevar a multas, demandas y otras acciones regulatorias, subrayando la importancia del cumplimiento de las regulaciones de protección de datos (Niraykumar, 2024).

El papel del aprendizaje automático en la mitigación de brechas

Con el aumento de la complejidad y el volumen de las amenazas cibernéticas, las tecnologías de aprendizaje automático (ML) han surgido como herramientas vitales en la ciberseguridad. El aprendizaje automático (ML) puede mejorar la detección de amenazas al analizar grandes cantidades de datos e identificar patrones indicativos de posibles brechas (Leveraging, 2025).

Esta capacidad es crucial para las organizaciones que luchan por gestionar el aumento de vulnerabilidades y ataques, ya que permite la identificación y respuesta proactiva a las amenazas emergentes (Wang, 2023).

Al aprovechar los algoritmos de aprendizaje automático, los equipos de seguridad pueden priorizar mejor las vulnerabilidades, detectar comportamientos inusuales y agilizar los esfuerzos de respuesta a incidentes, fortaleciendo en última instancia las defensas contra las brechas de ciberseguridad (Wang, 2023).

Material y métodos

Análisis Predictivo en Ciberseguridad

El análisis predictivo en ciberseguridad emplea técnicas de aprendizaje automático (ML) para prever posibles brechas de seguridad mediante el análisis de datos históricos e identificando patrones que preceden a tales incidentes. Este enfoque aprovecha los algoritmos para examinar grandes conjuntos de datos, incluidos los registros de tráfico de red, los patrones de comportamiento de los usuarios y las firmas de amenazas conocidas, facilitando el reconocimiento de anomalías indicativas de amenazas cibernéticas.

Mecanismos de Análisis Predictivo

Los modelos de aprendizaje automático están diseñados para aprender continuamente de los datos entrantes, lo que les permite adaptarse a las amenazas cibernéticas en evolución. Por ejemplo, el análisis predictivo puede analizar tendencias de violaciones de datos pasadas para identificar vectores de ataque comunes y activos vulnerables, estimando así la probabilidad de futuros ataques. Esta capacidad permite a las organizaciones fortalecer proactivamente sus defensas y mitigar riesgos antes de que ocurran las brechas.

Recolección de Datos y Ingeniería de Características

La efectividad de la analítica predictiva depende de la calidad y cantidad de los datos recopilados. El preprocesamiento de datos es esencial para eliminar el ruido y las inconsistencias, asegurando que la información alimentada a los modelos sea confiable.

Durante la extracción de características, se identifican atributos relevantes a partir de los datos, como los tamaños de los paquetes y las duraciones de las conexiones, que son críticos para mejorar la precisión y la eficiencia del modelo.

Algoritmos y sus Aplicaciones

Varios algoritmos de aprendizaje automático, incluyendo Random Forest, Árboles de Decisión y Redes Neuronales Convolucionales (CNNs), se emplean para análisis predictivo en ciberseguridad. Las CNN, en particular, han mostrado promesas en la detección de malware al aprender patrones complejos a partir de datos representados como imágenes, que los métodos tradicionales podrían pasar por alto.

Estos algoritmos clasifican las actividades como benignas o maliciosas, mejorando la velocidad y precisión de la detección de amenazas.

Limitaciones y Desafíos

La integración del aprendizaje automático (ML) en la ciberseguridad presenta varias limitaciones y desafíos que los investigadores y profesionales deben enfrentar.

Limitaciones de Datos

Una limitación significativa surge del alcance restringido de las fuentes de datos utilizadas en los estudios, favoreciendo predominantemente los materiales en inglés mientras se excluyen los recursos en otros idiomas. Esta limitación puede llevar a una comprensión sesgada del campo, ya que las contribuciones influyentes publicadas en otros idiomas podrían pasarse por alto. Además, la ausencia de una restricción temporal en las publicaciones podría resultar en la pérdida de valiosas perspectivas históricas que han dado forma a las prácticas y metodologías actuales en IA y ciberseguridad. Un enfoque equilibrado que abarque tanto la literatura contemporánea como la histórica es esencial para una comprensión integral del panorama de la investigación.

Escasez de Estudios de Caso del Mundo Real

La etapa temprana de la integración entre la IA y la ciberseguridad en contextos industriales también contribuye a la falta de estudios de caso sustanciales disponibles para revisión.

Esta escasez limita la capacidad de sacar conclusiones concretas o generalizar los hallazgos más allá de los marcos teóricos, obstaculizando un análisis más profundo de las implementaciones prácticas de la IA en la ciberseguridad. Para abordar esta brecha, se realizaron entrevistas con clientes de diversos sectores, integrando experiencias del mundo real para comprender mejor los desafíos prácticos que enfrentan las organizaciones en la evaluación de riesgos y la implementación de IA.

Problemas de Rendimiento del Modelo

Los modelos de aprendizaje automático en ciberseguridad enfrentan varios desafíos de rendimiento, incluyendo problemas de sobreajuste y subajuste. El sobreajuste ocurre cuando un modelo captura ruido de los datos de entrenamiento en lugar de los patrones subyacentes reales, lo que impacta negativamente su capacidad de generalización a datos no vistos.

Por el contrario, el subajuste surge cuando un modelo no puede aprender adecuadamente de los datos de entrenamiento, lo que resulta en altas tasas de error. Estos desafíos se ven agravados por los sustanciales recursos computacionales y de procesamiento de datos necesarios para una implementación efectiva de ML, lo que hace que la tecnología sea intensiva en recursos y potencialmente costosa para las organizaciones.

Vulnerabilidades Adversariales

La susceptibilidad de los modelos de aprendizaje automático a los ataques adversariales plantea otro desafío crítico. Los adversarios pueden manipular los datos de entrada para engañar a los algoritmos de ML, comprometiendo así la integridad y la fiabilidad de los modelos. Esto introduce una capa adicional de complejidad en el despliegue de sistemas de aprendizaje automático en ciberseguridad, lo que requiere estrategias robustas para fortalecer los modelos contra tales vulnerabilidades.

Gestión de Falsos Positivos

El equilibrio entre las tasas de verdaderos positivos y falsos positivos es crucial en las aplicaciones de ciberseguridad con ML. Si bien los verdaderos positivos son esenciales para una detección efectiva de amenazas, los falsos positivos pueden llevar al desperdicio de recursos mientras los equipos de seguridad persiguen amenazas inexistentes.

Gestionar los falsos positivos requiere una calibración cuidadosa de la sensibilidad y especificidad del modelo, lo que puede convertirse en una compensación crítica que afecte la eficiencia operativa y la productividad de los analistas. El ajuste continuo de los modelos de ML es necesario para mejorar la precisión y minimizar las tasas de falsos positivos con el tiempo.

Modelos Estadísticos

Para abordar el complejo desafío de predecir brechas de ciberseguridad, nuestro análisis se fundamentó en dos metodologías de aprendizaje supervisado de reconocida eficacia: el Bosque Aleatorio (Random Forest) y las Máquinas de Vectores de Soporte (Support Vector Machine).

Por un lado, el Bosque Aleatorio opera como un comité de expertos. En lugar de depender de un único árbol de decisión, este modelo construye un "bosque" compuesto por cientos de ellos. La clave de su éxito radica en que cada "árbol" se entrena con una muestra aleatoria y distinta de los datos.

Esta diversidad de perspectivas previene el "pensamiento grupal" (conocido como sobreajuste) y robustece la predicción final, que se obtiene por consenso. Una de sus ventajas más valiosas es su capacidad innata para cuantificar y clasificar la importancia de cada variable, revelando qué factores son los predictores más influyentes.

Por otro lado, la Máquina de Vectores de Soporte (SVM) adopta un enfoque más geométrico. Su objetivo es identificar la "frontera" que separe de la manera más clara y definitiva posible los dos grupos de datos: los sistemas que sufrieron una brecha y los que no. La genialidad del SVM reside en que no solo busca una frontera, sino aquella que maximiza el margen o la "zona de seguridad" entre ambas clases. Mediante el "truco del kernel", el modelo puede trazar estas fronteras de forma no lineal, adaptándose a relaciones intrincadas en los datos. Para cuantificar el riesgo, habilitamos su capacidad de generar una probabilidad asociada a cada predicción.

Ambos modelos fueron implementados para la tarea de clasificación binaria (predecir "brecha" o "no brecha"). Si bien ambos demostraron ser herramientas poderosas, en el contexto específico de nuestro análisis, el Bosque Aleatorio exhibió un rendimiento general superior, ofreciendo no solo predicciones precisas, sino también una valiosa interpretabilidad sobre los factores de riesgo.

Datos utilizados

Para llevar a cabo nuestro análisis, construimos un entorno de datos controlado pero profundamente realista. Este entorno se materializa en un conjunto de datos simulado que comprende 20,000 registros, donde cada uno representa un sistema de información empresarial hipotético.

De manera crucial, y en consonancia con la realidad, el conjunto de datos refleja que las brechas de seguridad son eventos excepcionales, no la norma. Así, la gran mayoría de los casos (un 84.2%) corresponden a sistemas sin incidentes, mientras que un 15.8% representa aquellos que sí sufrieron una brecha.

Para pintar un cuadro completo del panorama de seguridad de cada sistema, integramos una combinación de variables numéricas y categóricas. Incluimos métricas que capturan las amenazas técnicas, como el número de vulnerabilidades detectadas y su gravedad promedio en una escala de 0 a 10. Reconocimos el factor humano como un eslabón crítico, por lo que modelamos la tasa de clics en campañas de phishing. Finalmente, para contextualizar cada sistema, incorporamos variables categóricas que describen su frecuencia de actualización (desde diaria hasta nunca), el tipo de industria al que pertenece (como Finanzas o Tecnología) y su complejidad inherente (alta, media o baja).

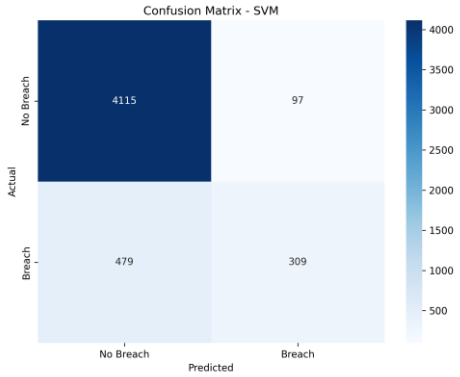
El corazón de nuestro análisis es la variable objetivo, Ocurrencia_Brecha, una variable binaria que responde a la pregunta fundamental: ¿sufrió el sistema una brecha de seguridad (1) o no (0)? Este diseño de datos, con su deliberado desequilibrio y su rica combinación de variables, nos proporciona un campo de pruebas robusto y fidedigno para entrenar y validar nuestros modelos predictivos en un escenario que emula los desafíos del mundo real.

Resultados

Al contrastar el rendimiento de nuestros dos enfoques metodológicos, observamos un panorama de alta competencia y matices reveladores. Ambos modelos, el Bosque Aleatorio y la Máquina de Vectores de Soporte, demostraron una capacidad predictiva notable, alcanzando una precisión global (accuracy) en torno al 88%. Sin embargo, un análisis más profundo revela las fortalezas distintivas de cada uno.

Figura1

Correlación variables

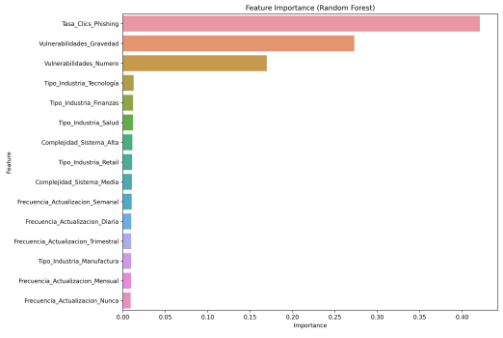


Fuente: Autores.

El modelo SVM se destacó por su agudeza al momento de identificar correctamente las brechas reales. Cuando este modelo levantaba una bandera roja, acertaba en un 76% de las ocasiones (precisión para la clase "Brecha"). Además, fue capaz de detectar el 78% del total de brechas ocurridas (recall). En contraparte, esta alta sensibilidad vino acompañada de una mayor variabilidad en sus resultados durante la validación cruzada, sugiriendo una ligera inestabilidad.

Figura 2

Caracteristicas importantes Random Forest



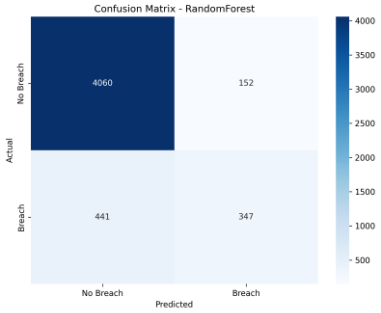
Fuente: Autores

El Bosque Aleatorio, por su parte, se presentó como un modelo de una consistencia excepcional. No solo mantuvo una precisión sólida, sino que mostró una desviación estándar mínima en las pruebas, lo que inspira una gran confianza en su fiabilidad. Aunque su precisión para detectar brechas fue ligeramente inferior a la del SVM (70%), su equilibrio general y su robustez lo convierten en un contendiente formidable.

Quizás el hallazgo más revelador de nuestro estudio no proviene de la comparación de modelos, sino del análisis de la importancia de las variables, una capacidad intrínseca del Bosque Aleatorio. Los resultados son contundentes: el factor humano, representado por la tasa de clics en phishing, se erige como el predictor más influyente por un margen considerable, acaparando un 42% del poder predictivo del modelo.

Figura 3

Matriz de correlación variables Random Forest



Fuente: Autores.

Le siguen, en orden de importancia, los factores técnicos: la gravedad de las vulnerabilidades (27%) y el número de vulnerabilidades (17%). Las variables contextuales, como el tipo de industria o la complejidad del sistema, aunque menos dominantes, actúan como piezas significativas que completan este rompecabezas predictivo.

Discusión

Los hallazgos de este estudio reafirman el potencial transformador del aprendizaje automático (ML) como herramienta para la predicción de brechas de ciberseguridad, una capacidad cada vez más crítica en el panorama de amenazas actual (Machhindra, 2023), (Jada, 2024).

La notable precisión global, cercana al 88%, alcanzada tanto por el modelo de Bosque Aleatorio (Random Forest) como por la Máquina de Vectores de Soporte (SVM), es consistente con la literatura que documenta la eficacia de estos algoritmos para la detección de anomalías y ataques (Dhaka, 2024). Sin embargo, más allá del rendimiento general, el análisis comparativo y de

importancia de variables ofrece matices cruciales para la implementación de estrategias defensivas informadas.

La evaluación del desempeño de los modelos revela un clásico dilema entre la agudeza diagnóstica y la estabilidad predictiva. La Máquina de Vectores de Soporte (SVM) demostró ser un detector altamente sensible, con una precisión del 76% y una exhaustividad (recall) del 78% para la clase "Brecha". En un contexto operativo, esto se traduce en un sistema de alerta temprana de gran valor: cuando el SVM identifica una amenaza, existe una alta probabilidad de que sea real. Esta capacidad de minimizar falsos negativos es fundamental en entornos de alto riesgo, donde pasar por alto una brecha real puede tener consecuencias catastróficas (Pujitha, et al, 2023).

No obstante, la mayor variabilidad observada durante la validación cruzada sugiere una posible susceptibilidad al sobreajuste o una sensibilidad a la composición específica de los datos de entrenamiento, lo que podría mermar la confianza en su rendimiento a largo plazo si no se realiza un monitoreo y reentrenamiento continuo (Srivastava, 2019).

En contraposición, el Bosque Aleatorio se perfiló como un modelo de robustez y fiabilidad excepcionales. Su consistencia, evidenciada por una desviación estándar mínima en las pruebas, lo convierte en una herramienta predecible y digna de confianza para el análisis de riesgos continuo (May, 2024). Aunque su precisión para la clase de interés (70%) fue marginalmente inferior a la del SVM, su equilibrio general y su resistencia a la variabilidad lo posicionan como un candidato ideal para una implementación a gran escala en sistemas de monitoreo de seguridad.

Esta estabilidad es un atributo clave, ya que fomenta la confianza de los analistas y reduce la "fatiga de alertas" que puede surgir de modelos más volátiles. La elección entre SVM y Bosque Aleatorio no es, por tanto, una cuestión de superioridad absoluta, sino una decisión estratégica que depende del apetito de riesgo de la organización y del objetivo específico de la predicción (Dhaka, 2024).

Quizás el hallazgo más revelador y de mayor impacto estratégico del estudio emana del análisis de la importancia de las variables, una ventaja inherente de los modelos basados en árboles como el Bosque Aleatorio. La conclusión es inequívoca: el factor humano, representado por la tasa de clics

en campañas de phishing, se erige como el predictor dominante, acaparando un 42% del poder predictivo.

Este resultado desplaza el foco tradicional de la ciberseguridad, a menudo centrado exclusivamente en defensas técnicas, y subraya que el eslabón humano sigue siendo el vector de ataque más explotado y, por ende, el indicador de riesgo más potente (Sing & Jha, 2021). Este hallazgo empírico refuerza la necesidad crítica de invertir en programas de concienciación y entrenamiento para empleados, no como una medida complementaria, sino como una línea de defensa central y medible.

Conclusiones

La conclusión principal de este estudio es doble y tiene profundas implicaciones para la estrategia de ciberseguridad moderna.

En primer lugar, se demuestra que no existe un modelo de aprendizaje automático universalmente superior, sino una complementariedad estratégica. El Bosque Aleatorio se establece como el pilar para un monitoreo de riesgo continuo y fiable, gracias a su robustez y consistencia. Por otro lado, la Máquina de Vectores de Soporte se revela como una herramienta de alta precisión, idónea para la investigación de alertas críticas donde minimizar los falsos negativos es la máxima prioridad. La elección, por tanto, no es una de exclusión, sino de asignación inteligente de recursos según el objetivo defensivo.

En segundo lugar, y de manera más contundente, el estudio concluye que el epicentro del riesgo de ciberseguridad reside inequívocamente en el factor humano. La abrumadora influencia predictiva de la tasa de clics en phishing (42%) relega a los factores puramente técnicos, como la gestión de vulnerabilidades, a un papel importante pero secundario.

Referencias bibliográficas

T. S., -, Theyjakshaya. D., & -, V. Vardhini. A. (2024). Cyber hacking breaches prediction and detection using machine learning. International Journal For Multidisciplinary Research, 6(3), 22653. https://doi.org/10.36948/ijfmr.2024.v06i03.22653

Department of Professional Security Studies, New Jersey City University, , Dhaka, Bangladesh. (2024). Comparative analysis of machine learning algorithms for predicting cybersecurity attack success: A performance evaluation. The American Journal of Engineering and Technology, 6(9), 81–91. https://doi.org/10.37547/tajet/Volume06Issue09-10

Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. Data and Information Management, 8(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Leveraging analytics to predict and prevent security breaches. (2025). Pavion. Retrieved July 11, 2025, from https://pavion.com/resource/leveraging-analytics-to-predict-and-prevent-security-breaches/

Machhindra, P. A., Vijay, B. N., Mahendra, B. S., Rahul, C. A., Anil, P. A., & Sunil, P. R. (2023). Enhancing cyber security through machine learning: A comprehensive analysis. 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), 1–6. https://doi.org/10.1109/ICCAKM58659.2023.10449547

May, R. (2024, September 30). Machine learning algorithms in cybersecurity. Ramsac Ltd. https://www.ramsac.com/blog/machine-learning-algorithms-in-cybersecurity/

Niravkumar Dhameliya. (2024). Machine learning in cybersecurity: A comprehensive analysis of intrusion detection systems. Journal of Sustainable Solutions, 1(4), 38–42. https://doi.org/10.36676/j.sust.sol.v1.i4.22

Pujitha, K., Nandini, G., Sree, K. V. T., Nandini, B., & Radhika, D. (2023). Cyber hacking breaches prediction and detection using machine learning. 2023 2nd International Conference on Vision

Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), 1–6. https://doi.org/10.1109/ViTECoN58111.2023.10157462

Raju, S. (2024). Adaptive security through machine learning with predictive approach to modern cyber threats. International Journal of Computer Applications, 186(50), 6–12. https://doi.org/10.5120/ijca2024924185

Singh, K., & Jha, S. (2021). Cyber threat analysis and prediction using machine learning. 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 1981–1985. https://doi.org/10.1109/ICAC3N53548.2021.9725445

Srivastava, T. (2019, August 6). 12 important model evaluation metrics for machine learning everyone should know(Updated 2025). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/

Susheela, S., Chandra, N. S., & Priyan, S. S. (2024). Predictive analytics-enabled cyber attack detection. International Journal of Innovative Science and Research Technology (IJISRT), 1242–1247. https://doi.org/10.38124/ijisrt/IJISRT24APR705

Tulsyan, R., Shukla, P., Singh, T., & Bhardwaj, A. (2024). Cyber security threat detection using machine learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, 08(10), 1–6. https://doi.org/10.55041/IJSREM37949

Wang, D. M. (2023, September 1). Machine learning in cybersecurity. Perspectives. https://www.paloaltonetworks.com/perspectives/the-future-of-machine-learning-in-cybersecurity/

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:

N/A

Nota:

El artículo no es producto de una publicación anterior.