Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), mayo-junio 2025,
Volumen 9, Número 3.

https://doi.org/10.37811/cl_rcm.v9i1

TÉCNICAS Y METODOLOGÍAS DE BIG DATA PARA EL ANÁLISIS DEL COMPORTAMIENTO DEL CIBERCRIMEN EN TIEMPOS DE PANDEMIA Y POSTPANDEMIA

TECHNIQUES AND METHODOLOGIES OF BIG DATA FOR THE ANALYSIS OF CYBERCRIME BEHAVIOR DURING AND AFTER THE PANDEMIC

Víctor Iván Valladarez González
Universidad Americana de Europa

Rodrigo Cadena Martínez
Universidad Americana de Europa

DOI: https://doi.org/10.37811/cl_rcm.v9i3.18235

Técnicas y Metodologías de Big Data para el Análisis del Comportamiento del Cibercrimen en Tiempos de Pandemia y Postpandemia

Víctor Iván Valladarez González[1]
ianvalladarez@hotmail.com
https://orcid.org/0009-0000-8509-8681
Universidad Americana de Europa

Rodrigo Cadena Martínez
rodrigo.cadena@aulagrupo.es
https://orcid.org/0000-0001-9323-6132
Universidad Americana de Europa

RESUMEN

La transformación digital acelerada por la pandemia de COVID-19 ha generado un incremento significativo en la actividad delictiva en entornos digitales. En este contexto, las técnicas y metodologías de Big Data han emergido como herramientas esenciales para el análisis y la predicción del comportamiento del cibercrimen. Este artículo presenta una revisión de las principales herramientas y enfoques de Big Data aplicadas al estudio del cibercrimen durante la pandemia y en la etapa postpandémica. Se destacan los métodos de minería de datos, análisis predictivo, aprendizaje automático y procesamiento de lenguaje natural, así como su contribución a la identificación de patrones delictivos, evaluación de riesgos y formulación de políticas públicas. El estudio concluye que el uso estratégico del Big Data no solo permite una mejor comprensión del fenómeno criminal digital, sino que también facilita respuestas más rápidas y eficaces por parte de las instituciones de ciberseguridad.

Palabras clave: big data, cibercrimen, pandemia, postpandemia, análisis predictivo

[1] Autor principal
Correspondencia: ianvalladarez@hotmail.com

Techniques and Methodologies of Big Data for the Analysis of Cybercrime Behavior During and After the Pandemic

ABSTRACT

The digital transformation accelerated by the COVID-19 pandemic has led to a significant increase in criminal activity within digital environments. In this context, Big Data techniques and methodologies have emerged as essential tools for the analysis and prediction of cybercrime behavior. This article presents a review of the main Big Data tools and approaches applied to the study of cybercrime during the pandemic and the post-pandemic period. It highlights methods such as data mining, predictive analytics, machine learning, and natural language processing, as well as their contributions to identifying criminal patterns, assessing risks, and shaping public policy. The study concludes that the strategic use of Big Data not only enables a better understanding of digital criminal phenomena but also facilitates faster and more effective responses by cybersecurity institutions.

Keywords: big data, cybercrime, pandemic, post-pandemic, predictive analysis

Artículo recibido 11 abril 2025
Aceptado para publicación: 15 mayo 2025

INTRODUCCIÓN

La pandemia de COVID-19 ha generado transformaciones profundas y aceleradas en el ecosistema digital global. La expansión del trabajo remoto, la virtualización de los servicios educativos y de salud, así como el crecimiento exponencial del comercio electrónico y las plataformas de interacción social, han reconfigurado los hábitos sociales y económicos, abriendo también nuevas oportunidades para actores maliciosos en el ciberespacio (Kshetri, 2021). Este cambio drástico en la estructura de interacciones digitales trajo consigo una alarmante escalada en los incidentes de cibercrimen, revelando la vulnerabilidad de los sistemas tecnológicos frente a amenazas cada vez más complejas y dinámicas.

En este escenario, el análisis del cibercrimen se ha convertido en una prioridad tanto para instituciones de seguridad como para organizaciones gubernamentales, privadas y académicas. Sin embargo, comprender el comportamiento delictivo en el ciberespacio exige herramientas capaces de procesar volúmenes masivos de datos, con velocidades y variedades que superan la capacidad de los métodos tradicionales de análisis criminal. Aquí es donde entra en juego el paradigma del Big Data, que permite la recopilación, procesamiento y análisis de datos estructurados y no estructurados, generados a gran velocidad y en enormes cantidades (Mayer-Schönberger & Cukier, 2013).

Big Data no solo transforma la forma en que se visualizan los fenómenos sociales, sino que también redefine las metodologías con las que se estudia el crimen. La capacidad de analizar datos de múltiples fuentes —como redes sociales, registros de tráfico de red, bases de datos judiciales, y foros en la dark web— permite trazar patrones de comportamiento criminal con mayor precisión y rapidez (Chen et al., 2018). En el contexto del cibercrimen, esta capacidad es crucial, ya que los atacantes pueden operar de manera anónima, transnacional y cambiante, haciendo que los enfoques estáticos pierdan eficacia.

Cibercrimen en contexto: pandemia y postpandemia

La pandemia generó una expansión sin precedentes en el uso de plataformas digitales. En 2020, la Organización Internacional del Trabajo (OIT) estimó que más del 30% de la fuerza laboral mundial operaba en modalidad remota. Este cambio, aunque beneficioso para la continuidad de las actividades económicas, expuso a millones de usuarios a entornos poco protegidos y sin medidas básicas de ciberseguridad (González & Méndez, 2022).

Las conexiones a redes domésticas, dispositivos personales sin protección adecuada, y la falta de cultura digital en muchos sectores propiciaron el terreno para un aumento explosivo de los delitos informáticos. Diversos estudios señalan que el número de ataques tipo ransomware creció más del 400% entre 2020 y 2021 (Check Point Research, 2021), mientras que los fraudes por phishing, ingeniería social y robo de identidad se multiplicaron en sectores como la banca digital, el e-commerce y los servicios de salud (Cisco, 2022). Durante la pandemia, ciberdelincuentes aprovecharon el miedo y la incertidumbre de la población, mediante campañas maliciosas relacionadas con supuestas curas, ayudas estatales o citas médicas falsas. Posteriormente, en la etapa postpandémica, estas dinámicas se han sofisticado, orientándose al robo de datos personales, extorsión y ataques a infraestructuras críticas.

Fundamentos del Big Data

El término Big Data hace referencia al tratamiento computacional de grandes volúmenes de datos que no pueden ser gestionados por sistemas tradicionales de bases de datos. Su definición está generalmente basada en las "cinco V": volumen, velocidad, variedad, veracidad y valor (Laney, 2001). El volumen alude a la cantidad masiva de datos generados cada segundo; la velocidad, a la rapidez con que estos datos son procesados; la variedad, a la diversidad de formatos (texto, imagen, audio, video, etc.); la veracidad, a la necesidad de evaluar su confiabilidad; y el valor, al conocimiento útil que se puede extraer.

El Big Data se apoya en un ecosistema tecnológico que incluye plataformas como Hadoop, Spark, y bases de datos NoSQL, junto con lenguajes de programación como Python y R. Además, el desarrollo de técnicas de machine learning, aprendizaje profundo (deep learning) y procesamiento de lenguaje natural (PLN) ha permitido que el análisis de datos no se limite a descripciones estáticas, sino que pueda generar modelos predictivos y prescriptivos útiles para la toma de decisiones en seguridad digital (Jordan & Mitchell, 2015).

Aplicación del Big Data al análisis del cibercrimen

En el campo de la ciberseguridad, el Big Data ha sido utilizado para múltiples propósitos: detección temprana de amenazas, identificación de patrones de ataque, clasificación de malware, análisis de comportamiento de usuarios, y monitoreo en tiempo real de redes.

Estas aplicaciones han revolucionado la manera en que los analistas detectan anomalías, responden a incidentes y predicen futuras acciones delictivas (Zhou et al., 2021).

Por ejemplo, mediante la minería de datos se pueden descubrir relaciones ocultas entre eventos aparentemente inconexos, como accesos no autorizados desde diferentes ubicaciones geográficas. El aprendizaje automático supervisado permite clasificar correos electrónicos como legítimos o de phishing, mientras que las redes neuronales profundas han sido aplicadas con éxito en la detección de bots y spam (Srinivas et al., 2020). Por su parte, el procesamiento de lenguaje natural facilita el análisis de amenazas emergentes en foros clandestinos, identificando intenciones delictivas antes de que se concreten ataques.

Teorías sobre el comportamiento delictivo en entornos digitales

Desde una perspectiva criminológica, el estudio del cibercrimen ha requerido la adaptación de teorías clásicas del delito a las particularidades del entorno digital. Una de las más relevantes es la teoría de las actividades rutinarias (Routine Activity Theory), propuesta por Cohen y Felson (1979), que sugiere que el crimen ocurre cuando coinciden un delincuente motivado, una víctima adecuada y la ausencia de un guardián. En el ciberespacio, estas condiciones son fácilmente replicables debido a la accesibilidad de las víctimas y la falta de vigilancia efectiva.

Otra teoría relevante es la teoría de la oportunidad racional (Rational Choice Theory), que plantea que los delincuentes sopesan riesgos y beneficios antes de actuar. En internet, el bajo riesgo de ser identificado, la posibilidad de actuar desde el anonimato y el alto beneficio económico (por ejemplo, en estafas con criptomonedas o datos personales) hacen que el cibercrimen sea percibido como una opción viable (Clarke & Cornish, 1985).

Además, la criminología computacional ha emergido como una subdisciplina que integra modelos matemáticos y simulaciones para estudiar el comportamiento delictivo en línea. La posibilidad de modelar agentes, escenarios y resultados a través de técnicas de simulación social abre nuevas rutas para entender y prevenir delitos cibernéticos (Malleson & Birkin, 2013).

Estado del arte: investigaciones recientes

Durante los últimos años, diversos estudios han demostrado el potencial del Big Data en el análisis del cibercrimen.

Por ejemplo, Han et al. (2022) utilizaron técnicas de análisis de sentimientos y clustering de texto para detectar amenazas cibernéticas emergentes en Twitter. Otro estudio de Diro & Chilamkurti (2018) desarrolló un modelo de detección de intrusiones basado en redes neuronales profundas que mejoraba significativamente la precisión de los sistemas de seguridad tradicionales.

Asimismo, organizaciones como la Europol y la Interpol han invertido en laboratorios de ciberinteligencia que utilizan Big Data para monitorear redes de cibercrimen organizado, rastrear la actividad de grupos de ransomware y prevenir ataques coordinados. Estas herramientas no solo permiten una respuesta más eficaz, sino también la identificación de tendencias delictivas, como el auge de las estafas de soporte técnico o el uso de inteligencia artificial por parte de los delincuentes.

Justificación y objetivos del estudio

Este artículo busca contribuir al cuerpo de conocimiento existente sobre la intersección entre Big Data y cibercrimen, analizando específicamente cómo las técnicas de análisis de datos han sido utilizadas —y pueden ser mejor aprovechadas— en contextos de crisis como la pandemia de COVID-19 y sus secuelas. Dada la velocidad con la que evoluciona el cibercrimen, es fundamental actualizar las metodologías de análisis para responder con eficacia a nuevas formas de ataque.

Entre los objetivos específicos se incluyen:

- Identificar las principales técnicas de Big Data aplicadas al análisis del cibercrimen.
- Analizar estudios de caso que demuestren la utilidad de estas técnicas durante la pandemia.
- Evaluar los retos éticos, técnicos y legales del uso de Big Data en este campo.
- Proponer recomendaciones para mejorar la capacidad de análisis y prevención del cibercrimen mediante estas tecnologías.

En síntesis, esta introducción establece las bases para una discusión profunda sobre la transformación del análisis del crimen digital a través de metodologías emergentes. Se parte del reconocimiento de un nuevo panorama delictivo, para el cual es necesario actualizar no solo las herramientas tecnológicas, sino también los marcos conceptuales y las estrategias institucionales.

METODOLOGÍA

Este artículo se enmarca dentro de un enfoque metodológico de tipo cualitativo-documental, mediante el cual se realizó una revisión sistemática de literatura científica y técnica sobre la aplicación de técnicas de Big Data al análisis del cibercrimen, específicamente en el contexto de la pandemia de COVID-19 y el periodo postpandémico. Se utilizó la estrategia PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) como guía para asegurar la transparencia, reproducibilidad y rigor del proceso de revisión (Moher et al., 2009).

Estrategia de búsqueda y selección de fuentes

La recopilación de información se realizó entre febrero y abril de 2025, a partir de bases de datos académicas reconocidas, tales como:

- Scopus
- Web of Science
- IEEE Xplore
- ScienceDirect
- Google Scholar

También se consultaron informes técnicos y documentos de organismos especializados como Europol, Interpol, ENISA (European Union Agency for Cybersecurity), Forbes Technology Council y empresas privadas del sector como Cisco, Check Point, Kaspersky, entre otras.

Se utilizaron palabras clave en inglés y español, combinando operadores booleanos para maximizar la precisión y cobertura de la búsqueda. Algunos de los descriptores empleados fueron:

"Big Data" AND "Cybercrime" AND "COVID-19",

"Cybersecurity" AND "Machine Learning" AND "Post-pandemic",

"Minería de datos" AND "Cibercrimen",

"Análisis predictivo" AND "pandemia",

"Tecnologías emergentes" AND "seguridad informática".

Criterios de inclusión y exclusión

Los criterios de inclusión fueron:

- Artículos científicos revisados por pares publicados entre 2018 y 2024.

- Documentos técnicos que incluyeran aplicaciones prácticas de Big Data en ciberseguridad.
- Estudios que analizaran el impacto de la pandemia en la evolución del cibercrimen.
- Fuentes que propusieran o analizaran modelos basados en aprendizaje automático, PLN, minería de datos o análisis de redes para detectar o prevenir delitos cibernéticos.

Criterios de exclusión

- Publicaciones duplicadas o con contenido redundante.
- Estudios sin sustento empírico o sin relación directa con el cibercrimen.
- Fuentes no verificables o de baja calidad metodológica.

Procedimiento de análisis

Una vez seleccionadas las fuentes, se procedió al análisis temático de los textos, clasificando los hallazgos en torno a cuatro categorías principales de análisis:

1. Técnicas de recolección y procesamiento de datos aplicadas al cibercrimen
2. Métodos de análisis predictivo y aprendizaje automático
3. Aplicaciones de minería de texto y procesamiento de lenguaje natural (PLN)
4. Desafíos éticos, legales y técnicos en el uso de Big Data en seguridad digital

Cada fuente fue codificada manualmente para identificar conceptos clave, enfoques metodológicos, contribuciones teóricas y resultados relevantes. Posteriormente, se sintetizaron los hallazgos en una tabla comparativa y se elaboró un análisis crítico de las principales aportaciones, detectando tanto patrones comunes como vacíos en la literatura.

Limitaciones del estudio

Si bien el presente estudio se realizó con criterios rigurosos de selección y análisis, es necesario reconocer ciertas limitaciones inherentes a los estudios de revisión:

- La rápida evolución del campo del Big Data y el cibercrimen puede generar que algunos aportes recientes no estén aún indexados en bases de datos formales.
- La literatura en idioma inglés predomina sobre otros idiomas, lo que puede haber excluido contribuciones relevantes en otras lenguas.
- No se incluyen estudios con datos primarios, ya que el alcance se limita a fuentes secundarias.

No obstante, se considera que los hallazgos aquí presentados brindan un panorama robusto, actualizado y crítico del estado del arte sobre las técnicas y metodologías de Big Data en la lucha contra el cibercrimen en entornos de pandemia.

RESULTADOS

La revisión sistemática permitió identificar tendencias clave, enfoques metodológicos recurrentes y desafíos emergentes en la aplicación del Big Data al estudio del cibercrimen en el contexto de la pandemia y postpandemia. A continuación, se presentan los hallazgos organizados en cuatro categorías de análisis.

Tabla 1. Resumen de Hallazgos por Categoría de Análisis

Categoría de Análisis	Técnicas / Enfoques Clave	Aplicaciones Relevantes	Herramientas Destacadas
Recolección y Procesamiento de Datos	Web scraping, logs, dark web, Hadoop, Spark, NoSQL	Monitoreo de amenazas, perfilamiento de riesgos	Maltego, Shodan, MongoDB
Análisis Predictivo y Machine Learning	Árboles de decisión, SVM, redes neuronales, clustering	Detección de intrusiones, clasificación de ataques	TensorFlow, Scikit-learn,Keras
Minería de Texto y Procesamiento de Lenguaje Natural (PLN)	TF-IDF, Word2Vec, LDA, BERT, análisis de sentimientos	Identificación de campañas de phishing, análisis de foros clandestinos	NLTK, SpaCy, HuggingFace
Desafíos Éticos, Legales y Técnicos	Privacidad,sesgos algorítmicos, explicabilidad, normativas de protección de datos (GDPR)	Evaluación de riesgos legales, diseño ético de sistemas de ciberseguridad	

Fuente: Elaboración propia.

Técnicas de recolección y procesamiento de datos aplicadas al cibercrimen

El primer paso en el uso de Big Data para el análisis del cibercrimen es la recolección eficiente de información proveniente de diversas fuentes. Durante la pandemia, los investigadores y entidades de ciberseguridad utilizaron técnicas como web scraping, sensores de red, logs de servidores, datos de plataformas sociales, y datos oscuros de la dark web para monitorear el comportamiento de potenciales cibercriminales (Choo, 2011; Sarker, 2021).

Una fuente crítica durante la pandemia fueron los sistemas de detección de eventos relacionados con COVID-19 en correos electrónicos y sitios web maliciosos. Herramientas como VirusTotal, Shodan y

Maltego fueron empleadas para trazar patrones de infección de malware y rastrear servidores de comando y control (Khan et al., 2022).

El procesamiento de estos datos implicó el uso de arquitecturas distribuidas como Apache Hadoop, Apache Spark y sistemas de almacenamiento NoSQL como MongoDB, capaces de soportar datos no estructurados provenientes de múltiples fuentes y de analizar flujos de datos en tiempo real (Katal et al., 2013). Estas infraestructuras permitieron no solo almacenar la información sino también realizar data cleaning, normalización y transformación para facilitar los análisis posteriores.

En cuanto a las fuentes utilizadas, los datos fueron extraídos de:

- Redes sociales (Twitter, Reddit, Telegram)
- Foros clandestinos (por ejemplo, 4chan, foros onion)
- Logs de firewalls y sistemas de detección de intrusiones (IDS/IPS)
- Bases de datos judiciales y policiales
- Bots de monitoreo de ciberamenazas

El uso coordinado de estas fuentes permitió construir perfiles de riesgo, identificar modus operandi recurrentes y anticipar eventos como campañas de phishing masivo.

Métodos de análisis predictivo y aprendizaje automático

Uno de los avances más significativos en el uso del Big Data para el estudio del cibercrimen ha sido la incorporación de modelos predictivos mediante algoritmos de machine learning (ML). Estos algoritmos permiten identificar patrones de comportamiento y predecir posibles amenazas antes de que se materialicen (Ahmed et al., 2016).

Entre los algoritmos más utilizados destacan

- Árboles de decisión (Decision Trees) y bosques aleatorios (Random Forests) para la clasificación de tipos de ataque.
- Máquinas de vectores de soporte (SVM) para la detección de anomalías en flujos de datos.
- Redes neuronales profundas (Deep Neural Networks), especialmente en el análisis de comportamiento en sistemas complejos.
- K-means y clustering jerárquico, usados para agrupar perfiles de amenazas o atacantes.

Por ejemplo, Diro y Chilamkurti (2018) implementaron un modelo basado en deep learning para la detección de intrusiones que superó en precisión a muchos de los sistemas tradicionales de detección. Durante la pandemia, estos modelos fueron aplicados para detectar accesos sospechosos a plataformas educativas y sanitarias, identificando intentos de brechas de seguridad en tiempo real.

Asimismo, los modelos de ML fueron entrenados con datasets públicos como CICIDS2017, UNSW-NB15, y CTU-13, que contienen registros de tráfico de red etiquetados con distintos tipos de ataques cibernéticos. En muchos casos, se utilizaron técnicas de oversampling y undersampling para balancear clases y mejorar la precisión de los modelos.

Además, los sistemas de análisis predictivo fueron integrados en SIEMs (Security Information and Event Management) que combinan datos históricos con detección en tiempo real, permitiendo respuestas automatizadas ante patrones delictivos detectados.

Aplicaciones de minería de texto y procesamiento de lenguaje natural (PLN)

El procesamiento de lenguaje natural (PLN) y la minería de texto han sido herramientas claves para analizar contenidos generados por usuarios en redes sociales, foros, mensajes de correo electrónico y otras plataformas donde se difunden amenazas cibernéticas o se reclutan víctimas (Camacho-Collados & Pilehvar, 2018).

Durante la pandemia, muchos actores maliciosos utilizaron Telegram, Discord y foros en la dark web para coordinar campañas de desinformación y compartir scripts maliciosos. Técnicas de PLN como la extracción de entidades nombradas (NER), análisis de sentimientos, topic modeling y detección de intenciones permitieron rastrear estas conversaciones y anticipar ataques como campañas de ransomware o fraudes financieros (Zhou et al., 2021).

Por ejemplo, mediante el análisis de palabras clave relacionadas con "ayuda gubernamental", "vacuna", "bono estatal" o "registro COVID", los investigadores lograron identificar correos de phishing y sitios web fraudulentos orientados a la recolección de datos personales.

Algunas técnicas destacadas incluyen:

- TF-IDF y Word2Vec para representar semánticamente textos maliciosos.
- LDA (Latent Dirichlet Allocation) para identificar temas dominantes en foros de cibercriminales.

- Modelos de lenguaje avanzados como BERT y GPT para comprender intenciones y detectar lenguaje fraudulento en tiempo real.

Estas herramientas han sido fundamentales no solo para investigar patrones de comunicación entre delincuentes, sino también para desarrollar sistemas de alerta temprana y alimentar dashboards de ciberinteligencia usados por gobiernos y fuerzas policiales.

Desafíos éticos, legales y técnicos en el uso de Big Data en seguridad digital

Si bien el uso de Big Data ha demostrado ser eficaz en la lucha contra el cibercrimen, también ha planteado serios desafíos éticos, legales y técnicos. El primero de ellos es la privacidad. Muchas de las técnicas de análisis requieren acceso a datos personales o metadatos que, aunque disponibles públicamente, pueden generar conflictos con los principios de protección de datos establecidos en normativas como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea (Zarsky, 2017).

Asimismo, la automatización de decisiones basada en algoritmos de aprendizaje automático plantea riesgos de sesgos algorítmicos, falsos positivos y profiling discriminatorio si no se diseña con criterios éticos y diversidad de datos. La opacidad de los modelos de deep learning, conocidos como "cajas negras", también ha generado debates sobre la explicabilidad y la rendición de cuentas en el contexto de la ciberseguridad (Burrell, 2016).

Desde el punto de vista técnico, el principal desafío es la gestión de datos no estructurados y la necesidad de integrar fuentes heterogéneas en tiempo real. La escalabilidad, latencia y confiabilidad de los sistemas también representan barreras para la implementación a gran escala de soluciones de Big Data en ciberinteligencia.

Por último, se identificó la necesidad de capacitación especializada. La implementación efectiva de estas tecnologías requiere perfiles interdisciplinarios, con competencias tanto en ciencia de datos como en derecho digital, criminología computacional y ética tecnológica.

DISCUSIÓN

Los resultados obtenidos evidencian que el uso de técnicas y metodologías de Big Data ha generado una transformación sustancial en la forma en que se analiza y combate el cibercrimen, especialmente en un contexto de crisis sanitaria y digitalización acelerada como el provocado por la pandemia de COVID-

19. Esta sección discute críticamente las implicancias de dichos hallazgos, conectándolos con la literatura existente y destacando sus impactos a nivel teórico, técnico y ético.

De la vigilancia reactiva al análisis proactivo

Tradicionalmente, las estrategias de ciberseguridad se han basado en una lógica reactiva, centrada en la respuesta a incidentes una vez ocurridos. Sin embargo, el uso de Big Data, junto con técnicas de inteligencia artificial, permite una transición hacia un modelo proactivo, donde las amenazas pueden anticiparse mediante la identificación temprana de patrones delictivos (Chen et al., 2018). Esta capacidad predictiva representa un cambio de paradigma, alineado con lo que algunos autores denominan "ciberinteligencia preventiva" (Bada & Nurse, 2019).

El análisis en tiempo real de grandes volúmenes de datos, provenientes de múltiples fuentes abiertas y cerradas, ha demostrado ser eficaz para detectar campañas maliciosas, dominios fraudulentos, y comportamientos anómalos en la red. No obstante, esta ventaja también plantea interrogantes sobre la sostenibilidad de estos modelos ante la constante evolución de las tácticas de los ciberdelincuentes, quienes, a su vez, comienzan a utilizar técnicas similares de automatización y aprendizaje automático para eludir la detección (Vincent et al., 2022).

Integración interdisciplinaria: ciencia de datos y criminología

El estudio del cibercrimen ya no puede abordarse exclusivamente desde una perspectiva técnico-informática. Como lo demuestran las aplicaciones de la criminología computacional, es necesario integrar enfoques sociológicos, jurídicos y éticos para comprender los factores que motivan, facilitan y regulan el comportamiento delictivo en entornos digitales (Malleson & Birkin, 2013). Las teorías de oportunidad, como la de las actividades rutinarias o la elección racional, siguen siendo relevantes, pero deben adaptarse a un espacio donde el tiempo, la geografía y la identidad se redefinen (Yar, 2013).

En este sentido, el uso de técnicas como el análisis de redes sociales (SNA) y el modelado de intenciones en texto no solo permite detectar delitos, sino también entender las dinámicas de grupo, los procesos de radicalización digital o la formación de mercados criminales en la dark web. Esta visión multidimensional resulta especialmente útil para el diseño de políticas públicas orientadas a la prevención del delito digital en el mediano y largo plazo.

Retos éticos y de gobernanza de los datos

Uno de los principales dilemas derivados del uso de Big Data en ciberseguridad es el equilibrio entre seguridad y privacidad. Si bien el análisis de datos masivos ha demostrado ser efectivo en la detección de amenazas, su aplicación sin criterios éticos puede derivar en violaciones a los derechos fundamentales, como el derecho a la intimidad o la protección de datos personales (Zarsky, 2017).

El uso de información sensible proveniente de redes sociales, correos electrónicos o foros debe considerar principios de consentimiento informado, minimización de datos y anonimización. Además, los sistemas de análisis automatizado deben ser auditables, explicables y libres de sesgos. Esto es especialmente relevante en contextos postpandémicos donde los gobiernos han incrementado sus capacidades de vigilancia bajo argumentos de emergencia sanitaria (Andrejevic, 2020).

La gobernanza de los algoritmos se convierte así en un aspecto crucial. Es necesario avanzar hacia marcos regulatorios que aseguren la transparencia, la rendición de cuentas y la protección de los derechos digitales, sin limitar el desarrollo tecnológico. Esto requiere una participación activa no solo de gobiernos y empresas, sino también de la sociedad civil, la academia y organismos internacionales.

Brechas técnicas y desigualdades globales

Otro aspecto a destacar es la desigual distribución de capacidades tecnológicas a nivel global. Mientras países desarrollados cuentan con infraestructuras de datos, talento humano y recursos económicos para implementar soluciones avanzadas de Big Data en ciberseguridad, muchos países en desarrollo enfrentan limitaciones significativas, tanto en conectividad como en formación técnica (UNCTAD, 2021).

Esta brecha digital no solo dificulta la implementación de sistemas de prevención del cibercrimen, sino que también convierte a estos países en objetivos más vulnerables para ataques de escala internacional. La cooperación regional y la transferencia de tecnología deben ser promovidas como estrategias de equidad y fortalecimiento de capacidades locales. De igual manera, es fundamental el desarrollo de marcos normativos armonizados para el tratamiento de evidencias digitales y la persecución penal transnacional.

Perspectivas futuras y sostenibilidad

A futuro, se prevé una expansión aún mayor del uso de Big Data en ciberseguridad, impulsada por avances en computación cuántica, blockchain y sistemas de inteligencia artificial general. Esta evolución plantea la necesidad de desarrollar sistemas más resilientes, interoperables y éticamente responsables.

En particular, la sostenibilidad de los modelos de análisis dependerá de su capacidad de adaptarse a nuevas formas de delito, como el uso de deepfakes para fraude o la manipulación algorítmica de resultados electorales. También será necesario fortalecer la alfabetización digital de la población y fomentar una cultura de la seguridad que no dependa exclusivamente de la tecnología, sino también de la conciencia colectiva sobre los riesgos del entorno digital.

CONCLUSIONES

El presente estudio ha permitido evidenciar el papel fundamental que desempeñan las técnicas y metodologías de Big Data en el análisis, comprensión y prevención del cibercrimen, particularmente en el marco de transformaciones digitales aceleradas por la pandemia de COVID-19 y sus efectos prolongados. A través de una revisión sistemática de literatura científica y técnica, se ha identificado que el uso de herramientas como la minería de datos, el aprendizaje automático, el procesamiento de lenguaje natural y los modelos predictivos han sido esenciales para anticipar amenazas, detectar patrones delictivos y apoyar la toma de decisiones estratégicas en ciberseguridad.

Asimismo, el estudio reafirma que el análisis del cibercrimen no puede entenderse únicamente desde una perspectiva tecnológica. Requiere una mirada integral que incorpore marcos teóricos provenientes de la criminología, la sociología, la ética digital y el derecho. Este enfoque interdisciplinario es crucial para interpretar los fenómenos emergentes del delito en entornos digitales, así como para diseñar políticas públicas más robustas, inclusivas y adaptadas a los retos del siglo XXI.

No obstante, también se identifican desafíos significativos. Entre ellos destacan las tensiones entre seguridad y privacidad, la necesidad de transparencia en el uso de algoritmos, la desigualdad tecnológica entre países, y la urgencia de establecer mecanismos de gobernanza de datos más efectivos. Estos factores condicionan la sostenibilidad y legitimidad de los sistemas de inteligencia basados en Big Data.

Finalmente, el estudio concluye que, aunque las herramientas de Big Data representan un avance sustantivo en la lucha contra el cibercrimen, su eficacia dependerá en gran medida de cómo sean implementadas, reguladas y evaluadas. La vigilancia ética, la formación de talento especializado y la cooperación internacional serán elementos clave para consolidar sistemas de ciberseguridad resilientes y respetuosos de los derechos humanos.

Recomendaciones

Fortalecer las capacidades institucionales en ciencia de datos aplicada a la ciberseguridad, promoviendo la capacitación interdisciplinaria de profesionales en análisis de datos, derecho digital, ética tecnológica y criminología computacional.

Impulsar marcos regulatorios que garanticen la protección de datos personales y la explicabilidad de los algoritmos, evitando sesgos discriminatorios y preservando los derechos fundamentales de los ciudadanos en entornos digitales.

Fomentar la cooperación internacional y regional en materia de inteligencia cibernética, compartiendo recursos tecnológicos, bases de datos y buenas prácticas para la prevención de delitos transnacionales en el ciberespacio.

Promover iniciativas de alfabetización digital y cultura de ciberseguridad dirigidas tanto a usuarios finales como a organizaciones públicas y privadas, con el objetivo de reducir la vulnerabilidad ante ataques.

Desarrollar centros de análisis de datos especializados en cibercrimen, que operen con una visión preventiva y que utilicen tecnologías emergentes de manera ética y responsable.

Invertir en sistemas de monitoreo y alerta temprana basados en Big Data, que permitan detectar tendencias delictivas en tiempo real y articular respuestas coordinadas entre múltiples actores.

Garantizar auditorías y evaluaciones periódicas de los sistemas basados en Big Data, para verificar su efectividad, detectar sesgos y asegurar que su uso se mantenga alineado con los principios de transparencia, equidad y rendición de cuentas.

REFERENCIAS BIBLIOGRAFICAS

Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

Andrejevic, M. (2020). Automated media. Routledge.

Bada, M., & Nurse, J. R. C. (2019). The social and psychological impact of cyber security breaches: A literature review. Research Report, University of Oxford.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12. https://doi.org/10.1177/2053951715622512

Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. Journal of Artificial Intelligence Research, 63, 743–788.

Check Point Research. (2021). Cyber Attack Trends: 2021 Mid-Year Report. Retrieved from https://research.checkpoint.com

Chen, H., Chiang, R., & Storey, V. (2018). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165–1188.

Choo, K.-K. R. (2011). The cyber threat landscape: Challenges and future research directions. Computers & Security, 30(8), 719–731. https://doi.org/10.1016/j.cose.2011.08.004

Cisco. (2022). Cybersecurity Threat Trends. Retrieved from https://www.cisco.com

Clarke, R. V., & Cornish, D. B. (1985). Modeling offenders' decisions: A framework for research and policy. In M. Tonry & N. Morris (Eds.), Crime and Justice (Vol. 6, pp. 147–185). University of Chicago Press.

Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761–768.

Europol. (2021). Internet Organised Crime Threat Assessment (IOCTA) 2021. Retrieved from https://www.europol.europa.eu

Fortinet. (2021). Global Threat Landscape Report. Retrieved from https://www.fortinet.com

González, L., & Méndez, R. (2022). Ciberseguridad y pandemia: análisis de las nuevas amenazas digitales. Revista de Seguridad y Defensa, 8(1), 45–62.

Han, S., Kim, J., & Lee, H. (2022). Real-time cyber threat detection using Twitter data: A hybrid approach combining sentiment analysis and topic modeling. Computers & Security, 112, 102512.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In 2013 Sixth International Conference on Contemporary Computing (pp. 404–409). IEEE. https://doi.org/10.1109/IC3.2013.6612229

Khan, M., Kiah, M. L. M., Madani, S. A., Khan, S. U., & Buyya, R. (2022). A survey on Big Data security and privacy issues. Journal of Cloud Computing, 11, 14.

Kshetri, N. (2021). COVID-19 and digital transformations in the global economy: The rise of cyber risks. Journal of Global Information Technology Management, 24(2), 121–132.

Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. META Group Research Note, 6(70), 1.

Malleson, N., & Birkin, M. (2013). Towards victim-oriented crime modelling in an agent-based simulation. Crime Prevention and Community Safety, 15(3), 188–203.

Mayer-Schönberger, V., & Cukier, K. (2013). Big Data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2(3), 160.

Srinivas, J., Das, A. K., & Kumar, N. (2020). Government regulations in cyber security: Framework, standards and recommendations. Future Generation Computer Systems, 100, 636–650.

UNCTAD. (2021). Digital Economy Report 2021: Cross-border data flows and development. United Nations Conference on Trade and Development.

Vincent, N., Hecht, B., & Sen, S. (2022). Algorithms ruin everything: Learning from the failures of algorithmic content moderation. ACM Transactions on Computer-Human Interaction (TOCHI), 29(1), 1–33.

Yar, M. (2013). Cybercrime and Society (2nd ed.). SAGE Publications.

Zarsky, T. Z. (2017). Incompatible: The GDPR in the age of big data. Seton Hall Law Review, 47(4), 995–1020.

Zhou, X., Han, L., & Zhang, Y. (2021). Cybercrime detection and prevention using NLP and deep learning. Journal of Information Security and Applications, 59, 102829.