-REVISTA-
Tecnopedagogía e Innovación

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Análisis de patrones y tendencias de las infracciones en ciberseguridad en un departamento de salud y servicios humanos

Analysis of Patterns and Trends in Cybersecurity Violations in a Department of Health and Human Services

• Fecha de recepción: 2023-06-01 • Fecha de aceptación: 2023-07-27 • Fecha de publicación: 2023-08-08

Jean Carlos Almeida[1], Joselyn Vergara Loor[2], Xavier Muñoz Pisco[3], Javier Guaña-Moya[4]
[1] Universidad Técnica Estatal de Quevedo, Quevedo, Ecuador
jean.almeida2015@uteq.edu.ec
ORCID: 0000-0003-0840-5689

[2] Universidad Técnica Estatal de Quevedo, Quevedo, Ecuador
joselyn.vergara2014@uteq.edu.ec
ORCID: 0000-0002-4544-7565

[3] Universidad Técnica Estatal de Quevedo, Quevedo, Ecuador
jeanny.munoz2016@uteq.edu.ec
ORCID: 0000-0002-0899-1625

[4] Instituto Superior Tecnológico Japón, Quito, Ecuador
eguana@itsjapon.edu.ec
ORCID: 0000-0003-4296-0299

Resumen

Este análisis de ciberseguridad enfocado en el sector de la salud destaca la prominencia de HCA Healthcare como líder en infracciones, subrayando la urgente necesidad de fortalecer las medidas de seguridad cibernética. La diversidad de vectores de ataque, con énfasis en los servidores de red, destaca la importancia crítica de abordar riesgos cibernéticos y debilidades en la gestión de documentos físicos. La distribución desigual de infracciones entre entidades resalta la necesidad de mejorar la ciberseguridad en proveedores de atención médica, líderes con 562 casos. La evolución temporal muestra un constante aumento de incidentes, llegando a 539 en 2023, enfatizando la necesidad de medidas sólidas de protección de datos. El análisis de tendencias destaca la prevalencia de "Hacking/IT" y acceso no autorizado, ofreciendo

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

perspectivas clave para estrategias proactivas de ciberseguridad. En resumen, este estudio técnico subraya la importancia crítica de mejorar la seguridad cibernética en el sector de la salud, abordando vectores de ataque específicos y tendencias emergentes para mitigar futuros riesgos.

Palabras clave: Ciberseguridad, clustering, análisis de patrones, evolución temporal.

Abstract

In this cybersecurity analysis focused on the healthcare sector, the prominence of HCA Healthcare as a leader in infractions is highlighted, underscoring the critical need to strengthen cybersecurity measures. The diversity of attack vectors, particularly on network servers, emphasizes the critical importance of addressing cybersecurity risks and vulnerabilities in the management of physical documents. The uneven distribution of infractions among entities underscores the urgency of improving cybersecurity in healthcare providers, leading with 562 cases. The temporal evolution reveals a continuous increase in incidents, reaching 539 in 2023, emphasizing the need for robust data protection measures. The analysis of temporal trends highlights the prevalence of "Hacking/IT" and unauthorized access, providing key insights for proactive cybersecurity strategies. In summary, this technical study underscores the critical importance of enhancing cybersecurity in the healthcare sector, addressing specific attack vectors, and emerging trends to mitigate future risks.

Keywords: Cybersecurity, Clustering, Pattern analysis, Temporal evolution.

Introducción

La misión del departamento de salud y servicios humanos de los Estados Unidos. Es mejorar la salud y el bienestar de todos los estadounidenses, brindando servicios humanos y de salud efectivos y fomentando avances sólidos y sostenidos en las ciencias subyacentes a la medicina, la salud pública y servicios sociales. En consonancia con este compromiso, la implementación de estrategias proactivas de ciberseguridad se vuelve fundamental para garantizar la integridad, confidencialidad y disponibilidad de la información relacionada con la salud, asegurando así un entorno seguro y protegido para el desarrollo continuo de avances en la atención médica y servicios sociales (Tsirintanis et al., 2023).

-REVISTA- Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

En la era contemporánea, los sistemas de información, el internet y la computación en la nube desempeñan un papel fundamental al respaldar el almacenamiento, la gestión y la utilización de datos tanto personales como organizacionales. Este entorno digital se presenta como un objetivo vulnerable para aquellos que buscan perpetrar acciones de robo, manipulación o perjuicio contra los dueños legítimos de la información. La creciente interconexión de dispositivos y la expansión de la digitalización han intensificado la exposición de estos activos, amplificando la necesidad imperante de implementar medidas de seguridad eficientes (Nicol, 2020), (Quezada & León, 2022).

Las compañías de seguridad cibernética y organizaciones privadas a nivel mundial establecen medidas y prevenciones de ataques y robo de información. Latinoamérica no escapa a la presencia de software malicioso, también conocido como malware; naciones como Brasil, Argentina, Uruguay, Chile, Colombia, Costa Rica, El Salvador, Guatemala, Honduras, México, Nicaragua, Paraguay, Perú y Venezuela han experimentado ataques de malware que comprometen tanto sistemas informáticos como información confidencial. En 2016, la empresa de seguridad informática ESET reveló que el 49 % de las pequeñas empresas y el 30 % de las medianas o grandes reportaron problemas relacionados con malware. El sector público se presenta como particularmente vulnerable, ya que carece de una política uniforme de identificación de riesgos, lo que dificulta la implementación de medidas efectivas de ciberseguridad (Chang, 2020).

El objetivo principal de los ciberataques es causar perjuicios financieros a las empresas, así como alcanzar otros objetivos, como los militares o políticos. Entre los métodos utilizados se encuentran adware, ataques de denegación de servicio distribuido (DDoS), doxing, gusanos, phishing, ransomware, spyware, troyanos y virus, entre otros. En respuesta, diversas organizaciones implementan varias soluciones de ciberseguridad para prevenir los daños ocasionados por estos ataques, aprovechando herramientas que monitorizan información en tiempo real acerca de las últimas tecnologías de la información (TI) (Guaña et al., 2022), (Moya, 2023).

La seguridad cibernética se ha convertido en una prioridad crítica en la era digital, especialmente para naciones como Estados Unidos, que han enfrentado diversos ataques,

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

planteando desafíos significativos para los derechos civiles. La interconexión de infraestructuras y sistemas de información ha dejado al país expuesto a amenazas cibernéticas avanzadas, que van desde intrusiones en redes gubernamentales hasta ataques dirigidos a entidades privadas. Estos incidentes no solo implican riesgos para la seguridad nacional, sino que también generan preguntas acerca de la privacidad y los derechos civiles de los ciudadanos. La delicada interacción entre la necesidad de defenderse contra amenazas cibernéticas y asegurar la preservación de los derechos individuales se ha convertido en un tema de gran importancia, requiriendo un equilibrio cuidadoso entre la seguridad y la salvaguardia de las libertades civiles en la era digital (Hernández et al., 2021).

La ciberseguridad se ocupa de desarrollar y aplicar medidas de protección para la información y la infraestructura tecnológica en el ámbito de las ciencias de la computación. Los piratas informáticos enfrentan ataques cotidianos, como el phishing o el malware, no solo en Ecuador, sino también en países con extensos sistemas empresariales. El auge tecnológico de la última década, impulsado por la nueva era digital y la globalización, ha propiciado un cambio sin precedentes en la evolución de la ciberseguridad (Cando & Chicaiza, 2021), (Hirare, 2017).

La industria de la ciberseguridad ha respondido de manera pronta al desafío actual mediante la investigación, desarrollo y aplicación del machine learning. Estas prácticas resultan cruciales para un análisis más profundo de las amenazas, mejorando la efectividad en la prevención de incidentes de seguridad. Ejemplos concretos de la implementación de inteligencia artificial en ciberseguridad incluyen la identificación de intrusos, la clasificación de malware, la detección de fraudes en tarjetas bancarias y la identificación de ataques de denegación de servicio (DDoS). En este contexto dinámico y en constante evolución, la integración de tecnologías avanzadas se ha vuelto esencial para anticipar y contrarrestar las amenazas cibernéticas. Este enfoque innovador no solo fortalece la seguridad digital, sino que también promueve la adaptabilidad necesaria para enfrentar los desafíos emergentes en el panorama (Fernández, 2018).

-REVISTA- Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

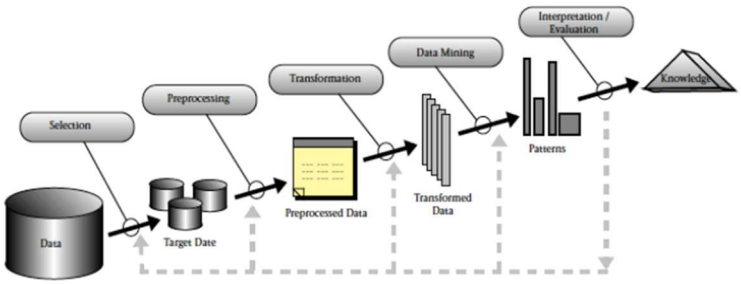
Metodología

Se llevó a cabo utilizando la base de datos obtenida de la oficina de derechos civiles del departamento de salud y servicios humanos de los Estados Unidos. Los datos recopilados de esta fuente van a servir como base para nuestro análisis de patrones y tendencias en las infracciones de ciberseguridad dentro del sector de la salud. Para el desarrollo de este trabajó se eligió la metodología KDD (Knowledge Discovery in Databases) siendo un proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y finalmente comprensibles en los datos siendo muy útil para el estudio en minería de datos.

Esta metodología consta de 5 etapas como se muestra en la figura 1, de esta manera, se caminó hacia la obtención de los correspondientes entregables basados en los resultados requeridos del proyecto (Timarán et al., 2016).

Figura 1.

Metodología KDD



Fuente: Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (Timarán et al., 2016).

Fase 1: Selección

Se llevará a cabo un procedimiento para identificar y obtener las fuentes de datos pertinentes para abordar la identificación de patrones y tendencias en relación con las violaciones de ciberseguridad en los datos de investigaciones de la oficina de derechos civiles de Estados Unidos donde se implementó una cadena de búsqueda para lograr obtener nuestra data considerando el siguiente parámetro de búsqueda como ***Ciberseguridad*** + ***"Ataque"*** + ***dataset***. Este proceso implica recopilar información sobre las infracciones que han ocurrido en el país, específicamente en Estados Unidos. También es fundamental tener en cuenta

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

factores contextuales. La elección cuidadosa de las variables asegurará que el proceso de descubrimiento de conocimiento se enfoque en los aspectos más significativos. Además, en esta fase inicial, se deben garantizar la precisión y la integridad de los datos recopilados, lo que exigirá la validación y verificación de la calidad de los datos.

La data obtenida presenta lo siguientes datos:

Tabla 1.

Descripción de variables del dataset

Variable	Descripción
Nombre de la entidad	Organización o entidad afectada por la infracción.
Estado	Estado o jurisdicción donde se produce la infracción.
Tipo de identidad cubierta	Tipo de información personal afectada
Individuos afectados	Número de personas cuyos datos se vieron comprometidos
Fecha de presentación de la infracción	Fecha en la que se informó por primera vez la violación de datos.
Tipo de incumplimiento	Naturaleza de la violación
Ubicación de la información violada	Donde se almacenaba la información comprometida (ej. servidores, bases de datos).
Socio comercial presente	Indica si un socio comercial estaba involucrado en la infracción de seguridad.

Fase 2: Preprocesamiento

Se llevará a cabo una exhaustiva limpieza y preparación de los datos, abordando valores faltantes, duplicados y posibles errores en los conjuntos de datos. Además, se implementarán técnicas de normalización para asegurar la coherencia y comparabilidad de los datos. Se prestará especial atención a la identificación y tratamiento de valores atípicos que puedan afectar la calidad de los análisis subsiguientes. La calidad y la integridad de los datos son fundamentales para asegurar la confiabilidad de los resultados en la minería de datos.

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

La data que se va utilizar esta en formato .csv como se muestra en la figura 2, en la cual se usara este formato para cumplir con todo el preprocesamiento de los datos.

Figura 2.

Data para análisis

Name of Covered Entity	State	Covered Entity Type	Individuals Affected	Breach Submission Date	Type of Breach	Location of Breached Informatior	siness Associate Pres	Web Description
Pan-American Life Insurance Group	Inc.	LA	Business Associate	94807	12/04/2023	Hacking/IT Incident	Network Server	Yes
Leggett & Platt Incorporated Employee Benefit Fund	MO	Health Plan	1200	12/04/2023	Hacking/IT Incident	Network Server	Yes	
Pan-American Life Insurance Group	Inc.	LA	Health Plan	105387	12/04/2023	Hacking/IT Incident	Network Server	No
EMS Management and Consultants Inc.	NC	Business Associate	2654	12/01/2023	horized Access/Disc	Paper/Films	Yes	
Community Healthcare Network	Inc.	NY	Healthcare Provider	500	11/30/2023	Hacking/IT Incident	Network Server	No
Fenway Community Health Center	Inc.	MA	Healthcare Provider	598	11/29/2023	Unauthorized Access/Disclosure	Paper/Films	Yes
Lakeview Healthcare System	LLC	FL	Healthcare Provider	2495	11/27/2023	Theft	Paper/Films	No
Foursquare Healthcare	Ltd.	TX	Healthcare Provider	10890	11/27/2023	Hacking/IT Incident	Network Server	No
Inline Plastics Corp.	CT	Health Plan	4853	11/27/2023	Hacking/IT Incident	Network Server	No	
Equality Health	LLC	AZ	Healthcare Provider	9240	11/27/2023	Hacking/IT Incident	Email	No
Oak Street Health	IL	Healthcare Provider	3238	11/24/2023	horized Access/Disc	Paper/Films	Yes	
Montrose Behavioral Health Hospital	Inc.	IL	Healthcare Provider	597	11/24/2023	Hacking/IT Incident	Email	No
Morrison Community Hospital District	IL	Healthcare Provider	122488	11/23/2023	Hacking/IT Incident	Network Server	No	
Molina Healthcare of Iowa	Inc.	IA	Business Associate	1647	11/22/2023	Hacking/IT Incident	Email	Yes
Saisystems International	Inc.	CT	Business Associate	10063	11/22/2023	Hacking/IT Incident	Network Server	Yes
Sierra County (Department of Public Health	nt of Behavior	CA	Healthcare Provider	2463	11/22/2023	Hacking/IT Incident	Email	No
East River Medical Imaging	PC	NY	Healthcare Provider	605809	11/22/2023	Hacking/IT Incident	Network Server	No
The Charles Lea Center	SC	Healthcare Provider	1250	11/22/2023	Hacking/IT Incident	Network Server	No	
Northwest Eye Care Professionals	OR	Healthcare Provider	950	11/22/2023	Hacking/IT Incident	Network Server	No	
U.S. Drug Mart	Inc.	TX	Healthcare Provider	13016	11/21/2023	Hacking/IT Incident	Network Server. Othe	Yes

La herramienta que se utiliza para la limpieza de datos es Python la cual es una herramienta de código abierto, además se usó la herramienta weka para encontrar los patrones y tendencias de las infracciones en ciberseguridad en un departamento de salud y servicios humanos.

En el proceso de limpieza de datos se validará los siguiente:

- **Campos Blancos**: Son espacios en los datos que carecen de información o contienen valores vacíos, requiriendo atención durante la limpieza para evitar afectar la integridad de los datos.
- **Campos Nulos:** Representan valores desconocidos o no asignados en los datos, y su manejo adecuado durante la limpieza incluye decisiones sobre cómo llenar esos valores faltantes para garantizar la fiabilidad de los resultados.

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 3.

Preprocesamiento de los datos

```
# Elimina las columnas especificadas
columnas_a_eliminar = [1,7, 8]  # Ten en cuenta que 6 representa la columna 7 y 7 representa la columna 8
datos = datos.drop(datos.columns[columnas_a_eliminar], axis=1)
datos
```

```
# Verifica si hay valores nulos en cada columna
datos = datos.isnull().sum()
datos
```

Fase 3: Transformación

Se llevó a cabo una serie de operaciones clave como considerar las fechas por separados, es decir; separar por año, fecha y meses para tener una mejor interpretación. Se procedió con la selección y proyección de atributos relevantes para el análisis, lo que implica la elección cuidadosa de variables que aporten valor al estudio de las infracciones. Además, se aplicarán técnicas de normalización y discretización para asegurar la consistencia y comparabilidad de los datos. Durante esta fase, también se considerará la posibilidad de derivar nuevas variables o características que puedan proporcionar perspectivas adicionales sobre los patrones emergentes.

Fase 4: Minería de datos

Para abordar la detección de patrones y tendencias sobre las infracciones en ciberseguridad en un departamento de salud y servicios humanos, se empleará algoritmos de minería de datos que nos permita cumplir con el objetivo, donde emplearemos el algoritmo de clustering (Ogunleye, 2021), (Bharadwaj et al., 2021), (Bokan & Santos, 2021).

Se pondrá énfasis en la identificación de variables clave relacionadas con ciberseguridad, tales como tipos de ataques y áreas vulnerables, para garantizar una representación completa de la complejidad del panorama de amenazas (Vishwakarma et al., 2023). Además, se incorporará una evaluación exhaustiva de los resultados obtenidos mediante el clustering, permitiendo una interpretación precisa de los agrupamientos identificados y proporcionando información valiosa para fortalecer las estrategias de seguridad cibernética en el departamento.

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Fase 5: Evaluación

Se lleva a cabo la evaluación de los patrones derivados de los modelos, donde se examina el comportamiento de las variables. Es importante resaltar que en esta etapa existe la posibilidad de retroceder a fases anteriores con el objetivo de desarrollar un nuevo modelo. Esta acción se emprende cuando se buscan comparaciones y, en este punto, se ha generado un conocimiento renovado.

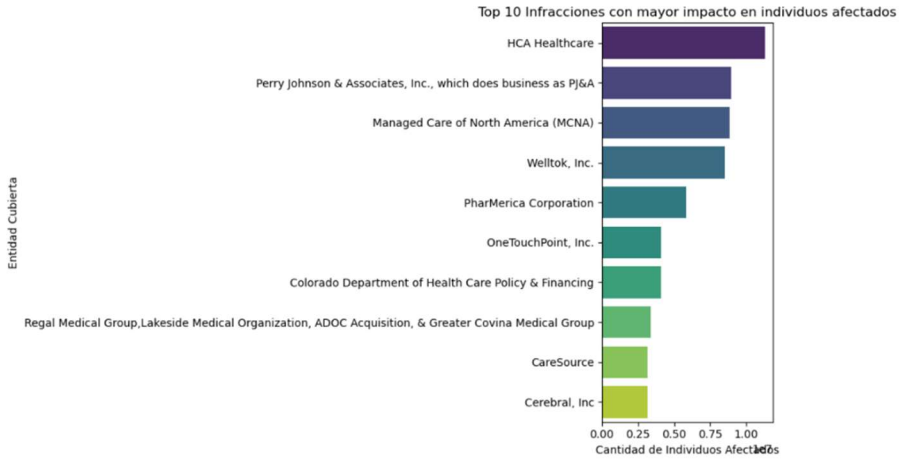
Resultado

Dentro de los resultados, tenemos como primer punto un análisis descriptivo en la figura 4, teniendo un gráfico de barra invertido como se indica en la figura donde la cantidad de individuos afectados por infracciones de seguridad en diferentes entidades, se observa que HCA Healthcare lidera la lista con un notable total de 11,270,000 individuos afectados, seguido por Perry Johnson & Associates, Inc. con 8,952,212 afectados. Este análisis destaca la magnitud significativa de la brecha de seguridad en el sector de la salud, evidenciando la importancia crítica de implementar medidas robustas de ciberseguridad. Es esencial que las entidades involucradas, especialmente las principales como HCA Healthcare, evalúen y fortalezcan sus protocolos de seguridad de la información para proteger la confidencialidad y privacidad de los datos de los pacientes. Este análisis proporciona una perspectiva clave para la toma de decisiones en la mejora continua de la seguridad cibernética en el ámbito de la atención médica, subrayando la necesidad de estrategias proactivas y soluciones innovadoras en la protección de datos sensibles de salud.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 4.

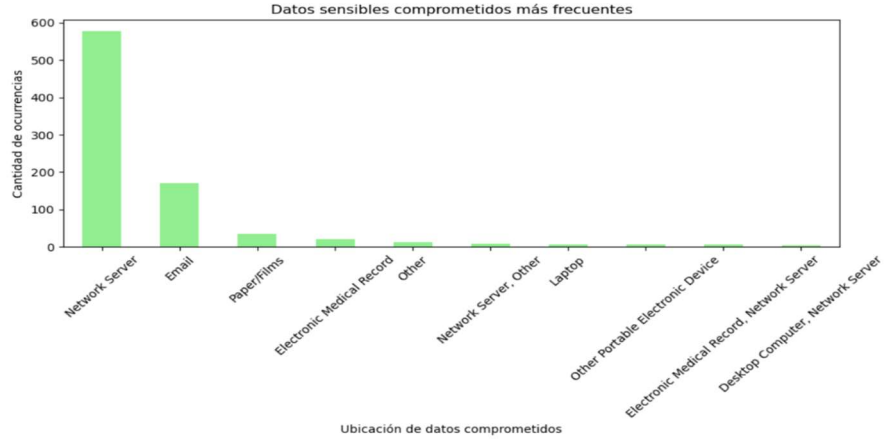
Infracciones con mayor impacto en individuos afectados



Tenemos un gráfico de barra como se indica en la figura 5 donde se revela que la mayoría de los incidentes, representados por un total de 578 casos, tienen su origen en servidores de red. Este hallazgo sugiere que las vulnerabilidades en las infraestructuras de red desempeñan un papel significativo en las brechas de seguridad. Las siguientes fuentes más comunes de incidentes incluyen el correo electrónico, con 170 casos, y documentos en papel o películas, con 35 casos. Estas cifras indican la diversidad de vectores de ataque utilizados, destacando la importancia de abordar tanto los riesgos cibernéticos como las potenciales debilidades en la gestión de documentos físicos.

Figura 5.

Datos sensibles más frecuentes

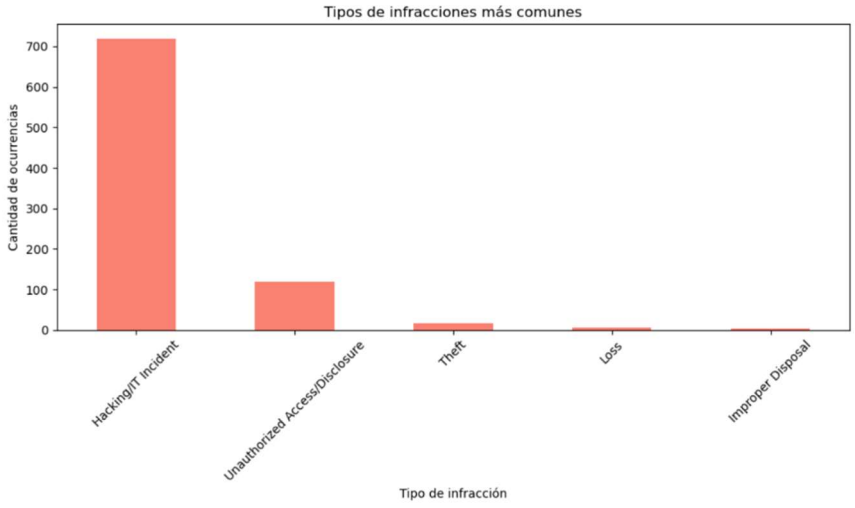


Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

En la figura 6, tenemos un gráfico de barra donde nos muestra que la principal causa de brechas de seguridad es "Hacking/IT Incident", con 719 casos, destacando la prevalencia de amenazas cibernéticas. En segundo lugar, "Unauthorized Access/Disclosure" con 119 casos destaca la importancia de gestionar adecuadamente los accesos no autorizados. Incidentes relacionados con "Theft" y "Loss" son menos frecuentes (16 y 6 casos respectivamente), subrayando la necesidad de abordar la seguridad física.

Figura 6.

Infracciones más comunes



El análisis de las infracciones en ciberseguridad revela una predominancia significativa de incidentes de hacking o tecnología de la información, representando el mayor número de ocurrencias, seguido por casos de acceso o divulgación no autorizados. Aunque menos comunes, el robo y la pérdida de datos también se registran. Sin embargo, la eliminación inadecuada de información muestra la menor frecuencia en el conjunto de datos. Este patrón destaca la urgencia de fortalecer las medidas de seguridad contra intrusiones y accesos no autorizados, además de resaltar la importancia de prácticas seguras para el manejo y disposición de datos sensibles, en línea con el objetivo de analizar patrones y tendencias en las infracciones de ciberseguridad.

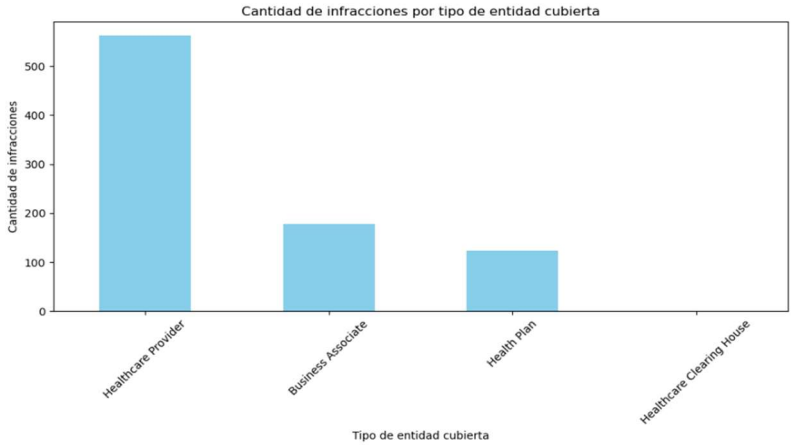
-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Cantidad de infracciones por tipo de entidad cubierta

Los resultados muestran una distribución desigual de infracciones de ciberseguridad entre entidades sanitarias en el periodo 2023. Los proveedores de atención médica registraron la mayor cantidad, con 562 infracciones, seguidos por asociados comerciales con 178 y planes de salud con 123. En contraste, las casas de compensación en salud reportaron solo una infracción. Estos números resaltan la necesidad urgente de mejorar las medidas de seguridad cibernética en los proveedores de atención médica y los asociados comerciales para salvaguardar la integridad de los datos y la privacidad de la información.

Figura 7.

Infracciones por entidad



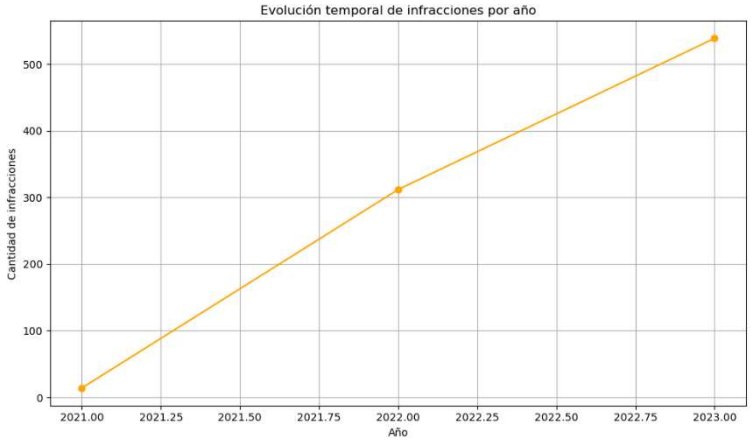
Evolución temporal de las infracciones

Los datos muestran un aumento significativo en la cantidad de infracciones de ciberseguridad a lo largo de los años. En el 2021 se registraron 14 infracciones, aumentando a 312 en el 2022 y alcanzando un total de 539 en el 2023. Este patrón revela un crecimiento constante y considerable en los incidentes de seguridad cibernética a lo largo de este periodo, destacando la urgente necesidad de implementar medidas más sólidas de protección de datos y seguridad informática para mitigar futuros riesgos.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 8.

Evolución temporal de las infracciones por año



Análisis de tendencias temporales por tipo de infracción

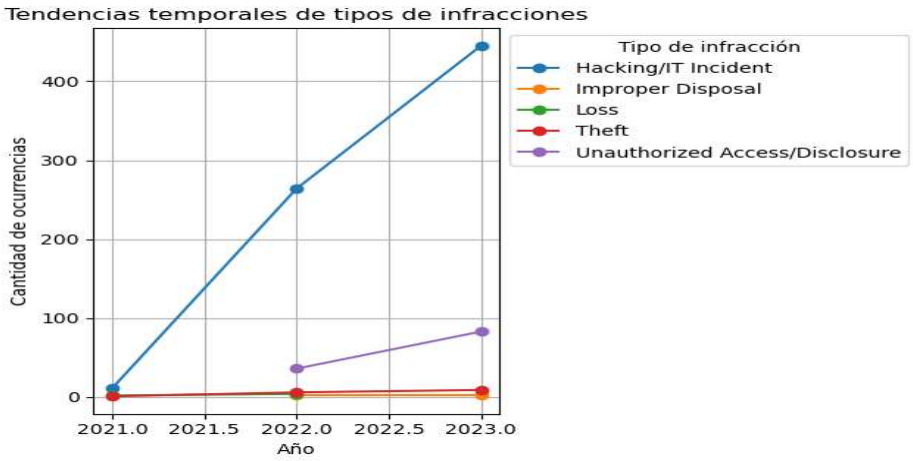
Estos datos detallan el tipo de brechas de seguridad reportadas en distintas categorías a lo largo de los años. En el 2021, se informaron 11 incidentes de hacking/IT, 2 de pérdida y 1 de robo, mientras que no se registraron casos de acceso/difusión no autorizados. En el 2022, los incidentes de hacking/IT aumentaron drásticamente a 264, con casos reportados de disposición inadecuada, pérdida, robo y acceso/difusión no autorizados. En el 2023, los incidentes de hacking/IT continúan siendo la categoría predominante con 445 casos, seguidos por 9 casos de robo, 83 de acceso/difusión no autorizados y una ausencia de datos sobre pérdida.

Este análisis detallado proporciona una perspectiva sobre la evolución de distintos tipos de brechas de seguridad a lo largo de los años, destacando la creciente incidencia de incidentes de hacking/IT y acceso no autorizado como áreas críticas a abordar en términos de fortalecimiento de la seguridad cibernética.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 9.

Tendencias temporales por tipo de infracción



Clustering

La distribución porcentual de las instancias en los clústeres muestra que el Clúster 2 es el más numeroso, representando el 40% del total de instancias. Le sigue el Clúster 1 con un 28%, luego el Clúster 3 con un 18%, y finalmente, el Clúster 0 con un 14%. Esta variación en la proporción sugiere que los clústeres no tienen una distribución uniforme y que algunos grupos pueden ser más representativos que otros en función de las características específicas del conjunto de datos.

Figura 10.

Distribución porcentual dentro de los clústeres

```
Clustered Instances

0       125 ( 14%)
1       243 ( 28%)
2       346 ( 40%)
3       156 ( 18%)
```

Se ha ejecutado el algoritmo SimpleKMeans en el conjunto de datos con una configuración específica. El algoritmo se ha inicializado con 4 clústeres (-N 4) y se han ajustado diversos parámetros, como el método de inicialización, el número máximo de candidatos, la poda periódica, la densidad mínima, los umbrales de distancia (t1 y t2), y el número máximo de iteraciones. Además, se ha utilizado la distancia euclidiana como métrica de similitud entre

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

las instancias. El conjunto de datos consta de 870 instancias con 7 atributos, que incluyen información sobre el tipo de entidad cubierta, el tipo de violación y la ubicación de la información comprometida. Se ha aplicado un filtro para ignorar ciertos atributos, como el nombre de la entidad cubierta, individuos afectados, fecha de presentación de la brecha y la presencia de un asociado comercial. La evaluación del modelo se ha realizado en el mismo conjunto de datos utilizado para entrenar el modelo (modo de prueba "evaluate on training data").

Figura 11.

Cantidad de interacciones

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A
Relation:     Examen-weka.filters.unsupervised.attribute.Remove-R2
Instances:    870
Attributes:   7
              Covered Entity Type
              Type of Breach
              Location of Breached Information
Ignored:
              Name of Covered Entity
              Individuals Affected
              Breach Submission Date
              Business Associate Present
Test mode:    evaluate on training data
```

El análisis revela que se aplicó un algoritmo de agrupamiento con un total de 3 iteraciones en un conjunto de datos relacionado con incidentes de brechas de seguridad. El objetivo principal de este proceso es agrupar los datos en clústeres significativos basados en la similitud de ciertas características. La métrica "Within cluster sum of squared errors" (Suma de errores cuadrados dentro de los clústeres) tiene un valor de 299.0, lo que sugiere la cantidad total de variabilidad dentro de los clústeres. Idealmente, un valor más bajo indicaría una agrupación más compacta y homogénea. Los puntos iniciales para cada clúster se seleccionaron aleatoriamente y se describen en términos de las características predominantes en cada grupo. Por ejemplo, el Cluster 0 se caracteriza por un 'Health Plan' afectado por un incidente de 'Hacking/IT' en un 'Network Server'. Este enfoque de inicio aleatorio influye en la formación de clústeres y afecta la convergencia del algoritmo.

-REVISTA-
Tecnopedagogía e Innovación

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 12.

Visualización de los grupos de clústeres

```
Number of iterations: 3
Within cluster sum of squared errors: 299.00000000000006

Initial starting points (random):

Cluster 0: 'Health Plan','Hacking/IT Incident','Network Server'
Cluster 1: 'Healthcare Provider','Hacking/IT Incident','Email  Other'
Cluster 2: 'Healthcare Provider','Hacking/IT Incident','Network Server'
Cluster 3: 'Business Associate','Hacking/IT Incident','Network Server'

Missing values globally replaced with mean/mode
```

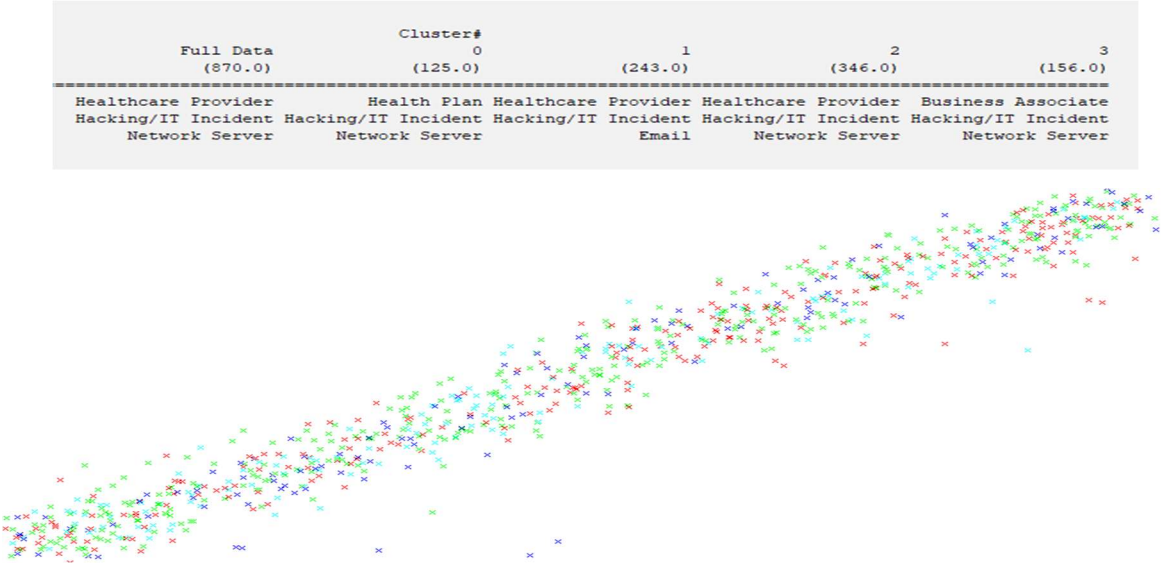
El atributo "Covered Entity Type" (Tipo de Entidad Cubierta) está representado predominantemente por proveedores de atención médica en los clústeres 0, 1 y 2, mientras que el clúster 3 está compuesto en gran medida por Asociados Comerciales. Esto sugiere que el algoritmo ha agrupado las entidades según el tipo de actor en el sector de la salud y la relación comercial. En cuanto al atributo "Type of Breach" (Tipo de Violación), todos los clústeres comparten la característica de incidentes de "Hacking/IT Incident", lo que indica una similitud en el tipo de amenaza experimentada por las entidades en cada clúster. El atributo "Location of Breached Information" (Ubicación de la Información Comprometida) revela patrones distintivos. El clúster 0 se destaca por incidentes que involucran una "Network Server", al igual que los clústeres 2 y 3.

En cambio, el clúster 1 está asociado principalmente con incidentes que afectan al "Email". Esta diferencia en la ubicación de la información comprometida puede sugerir variaciones en las tácticas de los incidentes de seguridad entre los clústeres.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Figura 13.

Clustering



Conclusiones

En el análisis descriptivo de las infracciones de seguridad, destaca la cifra significativa de individuos afectados en el sector de la salud, liderado por HCA Healthcare con 11,270,000 afectados. La diversidad de vectores de ataque, especialmente en servidores de red, resalta la necesidad de abordar riesgos cibernéticos y debilidades en la gestión de documentos físicos. La distribución de infracciones entre entidades destaca la urgencia de mejorar la ciberseguridad en proveedores de atención médica, líderes con 562 infracciones. La evolución temporal muestra un aumento constante de incidentes, alcanzando 539 en 2023, subrayando la necesidad de medidas sólidas de protección de datos. El análisis de tendencias temporales destaca la prevalencia de incidentes de "Hacking/IT" y acceso no autorizado. Estos resultados ofrecen perspectivas clave para estrategias proactivas de ciberseguridad, identificando áreas prioritarias como hacking y acceso no autorizado.

-REVISTA-
Tecnopedagogía e Innovación
Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Referencias

Bharadwaj, Prakash, K. B., & Kanagachidambaresan, G. R. (2021). Pattern recognition and machine learning. Programming with TensorFlow: Solution for Edge Computing Applications, 105-144.

Bokan, B., & Santos, J. (2021, April). Managing cybersecurity risk using threat based methodology for evaluation of cybersecurity architectures. In 2021 Systems and Information Engineering Design Symposium (SIEDS) (pp. 1-6). IEEE.

Cando-Segovia, M. R., & Chicaiza, R. P. M. (2021). Prevención en ciberseguridad: enfocada a los procesos de infraestructura tecnológica. 3 c TIC: cuadernos de desarrollo aplicados a las TIC, 10(1), 17-41.

Chang, J. E. A. (2020). Análisis de ataques cibernéticos hacia el Ecuador. Editora Adjunta, 2, 18.

Fernández Khatiboun, A. (2018). Machine Learning en Ciberseguridad.

Guaña-Moya, J., Sánchez-Zumba, A., Chérrez-Vintimilla, P., Chulde-Obando, L., Jaramillo-Flores, P., & Pillajo-Rea, C. (2022). Ataques informáticos más comunes en el mundo digitalizado. Revista Ibérica de Sistemas e Tecnologias de Informação, (E54), 87-100.

Hernández, E. F. T., Canizales, R. R., & Páez, A. V. (2021). La importancia de la ciberseguridad y los derechos humanos en el entorno virtual. Misión Jurídica, 14(20), 142-158.

Hirare, C. S. (2017). Ciberseguridad. Presentación del dossier. URVIO Revista Latinoamericana de Estudios de Seguridad, (20), 8-15.

Moya, J. G. (2023). La importancia de la seguridad informática en la educación digital: retos y soluciones. RECIMUNDO: Revista Científica de la Investigación y el Conocimiento, 7(1), 609-616.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55

Nicol, D. M. (2020). The Value of Useless Academic Research to the Cyberdefense of Critical Infrastructures. IEEE Security & Privacy, 18(01), 4-7.

Ogunleye, J. O. (2021). The Concept of Data Mining. In Data Mining-Concepts and Applications. IntechOpen.

Quezada Herrera, B. S., & León Yaguana, D. M. (2022). Revisión sistemática de la literatura relacionada con ciberseguridad apoyada con analisis de Big Data para actividades de red Team (Bachelor's thesis).

Timarán Pereira, S. R., Hernández Arteaga, I., Caicedo Zambrano, J., Hidalgo Troya, A., & Alvarado Pérez, J. C. (2016). Descubrimiento de patrones de desempeño académico.

Tsirintanis, K., Azzurro, E., Crocetta, F., Dimiza, M., Froglia, C., Gerovasileiou, V., ... & Katsanevakis, S. (2022). Bioinvasion impacts on biodiversity, ecosystem services, and human health in the Mediterranean Sea. Aquatic Invasions, 17(3), 308-352.

Vishwakarma, G. K., Paul, C., Hadi, A. S., & Elsawah, A. M. (2023). An automated robust algorithm for clustering multivariate data. Journal of Computational and Applied Mathematics, 429, 115219.

Período julio-diciembre 2023
Vol. 2 No. 2
rti@editorialscientificfuture.com
ISSN-L: 3028-8711 - **DOI:** https://doi.org/10.62465/rti.v2n2.2023.55