UNIVERSIDAD POLITÉCNICA SALESIANA

SEDE QUITO

CARRERA DE TELECOMUNICACIONES

DESARROLLO DE DETECCIÓN DE AMENAZAS CIBERNÉTICAS EN REDES EMPRESARIALES USANDO ALGORITMOS SUPERVISADOS

Trabajo de titulación previo a la obtención del
Título de Ingeniero en Telecomunicaciones

AUTORES: Alexander David Gutiérrez Espinel.
Alejandro Sebastián Granda Velastegui.
TUTOR: Juan Carlos Domínguez Ayala

Quito-Ecuador
2025

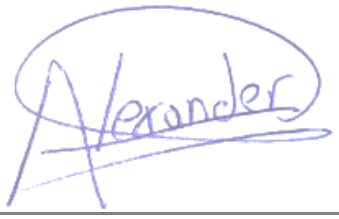
CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Alexander David Gutierrez Espinel con documento de identificación N° 0503130593 y Alejandro Sebastián Granda Velastegui con documento de identificación N° 1718525924; manifestamos que:

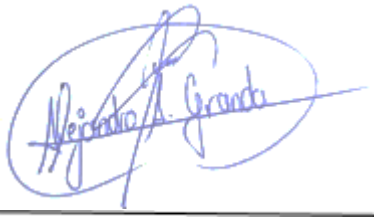
Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 22 de julio del año 2025

Atentamente,



Alexander David Gutiérrez Espinel
0503130593



Alejandro Sebastián Granda Velastegui
1718525924

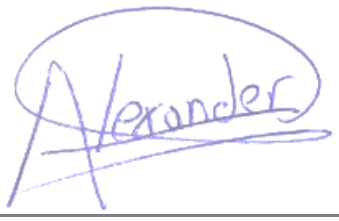
CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

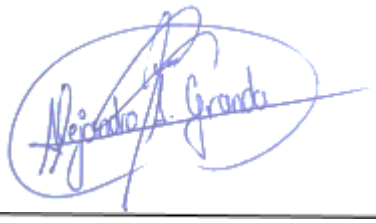
Nosotros, Alexander David Gutiérrez Espinel con documento de identificación N° 0503130593 y Alejandro Sebastián Granda Velastegui con documento de identificación N° 1718525924 expresamos nuestra voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos los autores del Artículo Académico: "Desarrollo de detección de amenazas cibernéticas en redes empresariales usando algoritmos supervisados", el cual ha sido desarrollado para optar por el título de: Ingeniero en Telecomunicaciones, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 22 de julio del año 2025

Atentamente,





Alexander David Gutiérrez Espinel
0503130593

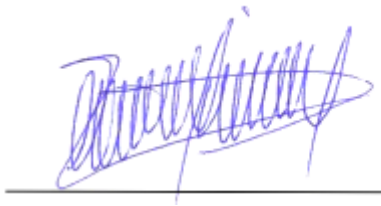
Alejandro Sebastián Granda Velastegui
1718525924

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Juan Carlos Domínguez Ayala con documento de identificación N° 1713195590, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: DESARROLLO DE DETECCIÓN DE AMENAZAS CIBERNETICAS EN REDES EMPRESARIALES USANDO ALGORITMOS SUPERVISADOS, realizado por Alexander David Gutiérrez Espinel con documento de identificación N° 0503130593 y por Alejandro Sebastián Granda Velastegui con documento de identificación N° 1718525924 obteniendo como resultado final el trabajo de titulación bajo la opción artículo académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 22 de julio del año 2025

Atentamente,



Ing. Juan Carlos Domínguez Ayala MSc.
171319559

DEDICATORIA

Dedico este articulo académico con todo mi cariño y profunda gratitud a Dios y a mis padres, Rosa Espinel y Armando Gutiérrez, quienes han sido un pilar fundamental en mi formación, pero especialmente a mi madre, Rosa, cuya entrega, apoyo incondicional y amor constante han sido la fuerza que más me ha sostenido a lo largo de este camino. Su ejemplo de sacrificio y perseverancia me ha inspirado en cada paso, y este logro no sería posible sin ella. Gracias por creer en mí siempre.

Alexander David Gutiérrez Espinel

El presente artículo académico se lo dedico principalmente a mi amada abuelita Mama Lucita, por ser la principal en motivarme a seguir una carrera universitaria, por guiarme y darme fuerza desde el cielo para no rendirme. A mis padres **Mónica Velastegui y a Jorge Granda**, por inculcarme los valores del respeto, la constancia y el amor por la familia, pilares fundamentales en cada uno de los logros que he alcanzado en mi vida. A mi querida esposa **Michelle Navarrete**, por ser mi compañera incondicional en cada paso de este camino, por su apoyo, comprensión y amor constante. A mi hijo **Sebitas**, mi mayor fuente de inspiración y motivo para esforzarme día a día. También dedico este trabajo a mis hermanos **Gabriela, Ricardo y Lorena**, por sus palabras de aliento, su compañía incondicional, por impulsarme siempre a seguir adelante y ser un ejemplo constante en mi vida. Gracias a cada uno de ellos, hoy soy una persona más fuerte, segura y comprometida con sus sueños.

Alejandro Sebastián Granda Velastegui

AGRADECIMIENTO

Agradezco profundamente a todas las personas que hicieron posible la realización de este artículo. A mi madre Rosa Espinel, por su amor incondicional, por ser mi mayor fortaleza y estar a mi lado en cada momento, tanto en los días más difíciles como en los más felices; su apoyo constante, sus palabras de aliento y su ejemplo de perseverancia han sido la base de todo lo que hoy he logrado. A mi padre Armando Gutiérrez, por su respaldo y confianza en cada paso de mi formación. A mi hermanita Renata, por ser una luz en mi vida, cuya ternura y alegría me inspiran cada día a seguir adelante. A mis abuelitos Mami Rocío y Papi Carlos que desde niño siempre me han brindado su amor y apoyo. A mi abuelito Papá Gonzalo por las historias inacabables que siempre llenaron mi vida de sabiduría, ternura y enseñanzas que aún resuenan en mi corazón, sus palabras contadas una y otra vez con cariño, me acompañaron en silencio durante este camino y siguen siendo un refugio al que siempre puedo volver. A mi compañero y más que compañero mi amigo Sebastián Granda, por ser ese apoyo constante, por su lealtad, compromiso y por siempre estar en las buenas y en las malas. A mi novia Ambar, por acompañarme en mi carrera universitaria por su comprensión, cariño y festejar mis logros como suyos. A mis docentes, por su guía y dedicación a lo largo de mi carrera, y especialmente al Ing. Juan Carlos Domínguez, por su valiosa orientación y acompañamiento durante el desarrollo de este trabajo. Agradezco a mis compañeros que han sido testigos de este proceso y de tantas experiencias que las llevare en mi corazón.

Alexander David Gutiérrez Espinel

AGRADECIMIENTO

Quiero agradecer, en primer lugar, a Dios, por ser mi guía constante, por brindarme fortaleza, sabiduría y fe en cada etapa de este camino académico. A mis padres, Mónica Velastegui y Jorge Granda, por su amor incondicional, sus enseñanzas y su apoyo firme en todo momento. Este logro también les pertenece.A mi esposa, Michelle Navarrete, por ser mi soporte emocional y compañera fiel en este proceso; y a mi hijo Sebitas, por ser mi mayor inspiración, la luz que me impulsa a seguir adelante con determinación y esperanza.Agradezco profundamente a mis hermanos, Gabriela, Ricardo y Lorena, por estar siempre presentes con sus palabras de aliento y motivación.A mi tutor de tesis, Juan Carlos Domínguez, por su guía, paciencia y compromiso durante el desarrollo de este trabajo. Su experiencia y consejos fueron fundamentales para alcanzar los objetivos propuestos.A mi compañero de tesis, Alexander Gutiérrez, por su dedicación, trabajo en equipo y apoyo constante. Compartir este reto contigo hizo que el camino sea más productivo y gratificante.Extiendo también mi gratitud a mis profesores, por su dedicación, exigencia y compromiso con la enseñanza. Sus conocimientos y ejemplo han sido fundamentales en mi formación profesional y personal.Finalmente, a mis compañeros de carrera, gracias por su amistad, colaboración y por todos los momentos compartidos. Cada experiencia vivida ha enriquecido esta etapa, dejándome recuerdos imborrables y valiosas lecciones de vida.

Alejandro Sebastián Granda Velastegui

DESARROLLO DE DETECCIÓN DE AMENAZAS CIBERNÉTICAS EN REDES EMPRESARIALES USANDO ALGORITMOS SUPERVISADOS

Sebastián Granda[1], Alexander Gutiérrez[2], and Juan Carlos Domínguez[3]

[1] Universidad Politécnica Salesiana, Quito, Ecuador
agrandav1@est.ups.edu.ec
https://www.ups.edu.ec/
[2] Universidad Politécnica Salesiana, Quito, Ecuador
agutierreze@est.ups.edu.ec
https://www.ups.edu.ec/
[3] Universidad Politécnica Salesiana, Quito, Ecuador
jdominguez@ups.edu.ec
https://www.ups.edu.ec/

Resumen

En la actualidad, ransomware manifiesta una seria amenaza para la operación de las empresas, al detener procesos internos y servicios esenciales. Frente al cambio constante de estas técnicas, se vuelve indispensable contar con métodos de detección más eficaces y flexibles. Esta dificultad, se aplicó una estrategia basada en el modelo `CRISP-DM` combinando algoritmos de aprendizaje automático con análisis de acceso al almacenamiento. El estudio utilizó el conjunto de datos público `RanSAP`, que incluye registros de actividad tanto de programas legítimos como de variantes de ransomware. Mediante la obtención de características y el uso de técnicas como SHAP, fue posible analizar el comportamiento del sistema. Entre los modelos entrenados, Random Forest brindó mejores resultados, con un F1-score de 0.96 y un AUC de 0.98, distinguiéndose como la opción más capaz.

Palabras clave

Aprendizaje automático, ransomware, CRISP-DM, interpretabilidad, SHAP, detección de amenazas.

Abstract

Currently, ransomware poses a serious threat to business operations by halting internal processes and essential services. Given the constant evolution of these techniques, more effective and flexible detection methods are essential. To address this difficulty, a strategy based on the CRISP-DM model was applied, combining machine learning algorithms with storage access analysis. The study used the public RanSAP dataset, which includes activity logs of both legitimate programs and ransomware variants. By extracting features and using techniques

such as SHAP, it was possible to analyze system behavior. Among the trained models, Random Forest provided the best results, with an F1 score of 0.96 and an AUC of 0.98, standing out as the most capable option.

Keywords

Machine learning, ransomware, CRISP-DM, interpretability, SHAP, threat detection

1. Introducción

La tecnología y su rápido avance ha cambiado la forma de operación de las empresas, simplificando procesos, optimizando el acceso de servicios y mejorando la conectividad. Sin embargo, la creciente digitalización también ha extendido el área de ataque, dejando a las organizaciones más expuestas a amenazas como el ransomware. Esta clase de malware cifra la información del sistema y solicita realizar un pago económico a cambio de su liberación. De acuerdo con [3], este tipo de ataque ha ganado fuerza debido a su impacto directo sobre los activos digitales y su alta tasa de éxito [3]. Durante 2024, se estima que más del 60 % de las empresas a nivel global se vieron afectadas por ransomware, provocando interrupciones operativas, pérdidas financieras y daño reputacional [1].

Las estrategias de defensa convencionales, como el uso de firmas o reglas estáticas, han perdido efectividad frente a nuevas variantes de ransomware que incorporan técnicas avanzadas de evasión [4]. Por lo tanto, para hacer frente a esta situación, se han generado perspectivas basadas en inteligencia artificial, en especial el aprendizaje automático, como opciones prometedoras para detectar comportamientos anómalos antes de que el daño sea irreparable [15]. Sin embargo, la falta de explicabilidad en los modelos de IA representa un desafío, especialmente en redes empresariales complejas, donde coexisten múltiples procesos, usuarios y servicios críticos que requieren decisiones justificables. Herramientas como **SHAP (SHapley Additive Explanations)** permiten descomponer las predicciones del modelo para identificar qué características influyen en cada decisión, facilitando así su adopción al proporcionar explicaciones comprensibles para los analistas [5].

En general este trabajo presenta una metodología sistemática para la detección de ransomware, alineada con CRISP-DM y adaptada en el análisis de patrones de acceso al almacenamiento. A diferencia de estudios basados en el tráfico de red o eventos del sistema, la propuesta actual se basa en el conjunto de datos RanSAP, que recopila más de 200 ejecuciones de ransomware y software legítimo bajo condiciones controladas [5]. Estas ejecuciones fueron capturadas mediante hipervisores ligeros, lo que permite registrar trazas fiables sin interferir en el comportamiento original de los programas [7].

La metodología incluye la transformación de los registros en atributos agregados, como el throughput de escritura (Twrite), la varianza de direcciones (Vwrite) y la entropía media de bloques (Hwrite), métricas que han demostrado ser útiles para distinguir patrones de comportamiento [13]. Con estas variables, se entrenaron tres algoritmos supervisados: Random Forest, Support Vector Machine (SVM) y Gradient Boosting. Estos modelos han sido ampliamente utilizados por su eficacia en clasificación, robustez frente a datos ruidosos y capacidad de adaptación [6]. Además, se aplicó SHAP para interpretar el aporte de cada característica a las decisiones del modelo.

Aunque el ransomware ha sido ampliamente abordado en investigaciones previas, persisten importantes desafíos prácticos, especialmente en entornos empresariales reales donde los modelos deben ser eficaces, adaptables e interpretables. Este estudio contribuye al área al proponer un enfoque replicable que prioriza la explicabilidad del modelo, aspecto crucial ante la constante evolución de las amenazas y la necesidad de cumplimiento normativo en sectores corporativos.

1.1. Contribuciones de este artículo

Este trabajo presenta un enfoque integral para la detección de ransomware en entornos empresariales, fundamentado en el análisis de patrones de acceso al almacenamiento mediante aprendizaje automático. A diferencia de otros estudios centrados en eventos del sistema o tráfico de red, se emplea el conjunto de datos RanSAP, que permite observar el comportamiento dinámico de múltiples familias de ransomware y aplicaciones benignas en condiciones controladas[7]. Las principales contribuciones del artículo son las siguientes:

- Se desarrolla un pipeline reproducible para la detección de ransomware a partir de métricas dinámicas de bajo nivel, obtenidas sin interferencia del entorno, gracias al uso de hipervisores ligeros.
- Se valida experimentalmente el valor del dataset RanSAP para la clasificación de amenazas, evidenciando que variables como el throughput de escritura, la entropía y la varianza de acceso permiten distinguir eficazmente entre actividades legítimas y maliciosas.
- Se comparan tres algoritmos supervisados (Random Forest, SVM y Gradient Boosting), destacando la importancia de la interpretabilidad en ciberseguridad, apoyada en técnicas como SHAP [5].
- Se incluyen visualizaciones adicionales como histogramas, boxplots y mapas de calor de correlación, así como proyecciones con t-SNE. Estas herramientas permiten explorar las diferencias estadísticas entre clases, identificar relaciones entre variables y facilitar el análisis replicable por parte de otros investigadores.

1.2. Relación con estudios previos

Diversos estudios han abordado la detección de ransomware mediante el análisis de patrones de acceso al almacenamiento, siendo Hirano y Kobayashi [7] referentes en el uso de aprendizaje automático sobre métricas recolectadas con hipervisores ligeros. Sin embargo, sus experimentos estaban limitados en variedad y tamaño de muestra. El presente trabajo utiliza el conjunto datos público

`RanSAP` [7], que amplía significativamente la diversidad de familias de ransomware, aplicaciones benignas y condiciones experimentales, permitiendo una evaluación más robusta y generalizable de los modelos propuestos. Esto contribuye a la literatura al demostrar el valor de patrones de acceso a bajo nivel y destaca la importancia de emplear conjuntos de datos abiertos y variados para fortalecer la reproducibilidad en ciberseguridad [8].

2. Resumen del conjunto de datos RanSAP

El conjunto de datos *RanSAP* representa la base experimental de esta investigación. Fue construido específicamente para analizar la actividad de ransomware y software legítimo a través de sus interacciones con el subsistema de almacenamiento. En total, se recopilaron más de 200 ejecuciones controladas, correspondientes a 15 familias de ransomware y 9 aplicaciones benignas, lo que se traduce en millones de operaciones individuales de lectura y escritura registradas en archivos CSV. Cada registro incluye atributos como marca de tiempo, dirección lógica de bloque (LBA), tamaño de acceso y nivel de entropía, lo que permite capturar el comportamiento a muy bajo nivel.

Las trazas fueron recolectadas utilizando hipervisores ligeros para minimizar la alteración del entorno y garantizar la validez de los datos. En comparación con estudios anteriores como el de [7], que utilizaron conjuntos más reducidos y menos diversos, *RanSAP* ofrece mayor riqueza experimental, tanto en volumen como en variedad. Esta amplitud de escenarios fortalece la robustez del análisis y mejora la capacidad de los modelos para generalizar en tareas de detección.

2.1. Estructura y formato

Los datos recopilados en RanSAP se encuentran organizados en archivos CSV, donde cada archivo representa una ejecución independiente de un software específico, ya sea una familia de ransomware o una aplicación legítima. Cada uno de estos archivos fue generado bajo condiciones experimentales controladas, lo que garantiza la consistencia entre las ejecuciones. En cada archivo, las filas corresponden a operaciones individuales de lectura o escritura realizadas por el programa observado.

La Tabla 1 muestra un extracto del dataset original, donde cada registro incluye variables como: la marca de tiempo, la dirección lógica de bloque (LBA), el tamaño de acceso en bytes y el nivel de entropía del bloque afectado. Estas características permiten capturar patrones a bajo nivel sobre cómo interactúa el software con el sistema de almacenamiento.

Tabla 1. Registros del dataset original.

ts_s	ts_ms	lba	size	entropy	family	label
1596701325	634204541	6587184	4096	0.270780	Cerber	1
1596701325	63566537	12376496	4096	0.180318	Cerber	1
1596701325	63762813	589688	4096	0.154709	Cerber	1
1596701325	64521952	6438944	4096	0.379392	Cerber	1
1596701325	64628228	2254104	4096	0.265534	Cerber	1

Descripción de atributos:

- **ts_s**: Momento en que ocurrió la operación (segundos desde época UNIX).
- **ts_ms**: Precisión adicional de la marca de tiempo en milisegundos.
- **lba**: Dirección lógica del bloque accedido (lectura o escritura).
- **size**: Tamaño del bloque accedido, expresado en bytes.
- **entropy**: Nivel de aleatoriedad en el contenido del bloque.
- **family**: Familia de ransomware o nombre del software benigno.
- **label**: Clasificación binaria (1 = ransomware, 0 = benigno).

Posteriormente, se aplicó una fase de ingeniería de características sobre ventanas temporales para extraer variables agregadas que capturen el comportamiento estadístico del acceso al almacenamiento. Estas nuevas variables fueron utilizadas como entradas para los modelos de clasificación.

A continuación, la Tabla 2.

Tabla 2. Registros tras la ingeniería de características.

Twrite	Vwrite	Hwrite	label	family
1426892.8	4.413665e+14	0.329781	0	SDelete
1596108.8	3.827644e+14	0.336918	0	SDelete
1508044.8	3.948839e+14	0.338263	0	SDelete
1453158.4	3.627244e+14	0.342738	0	SDelete
1455206.4	3.203119e+14	0.347195	0	SDelete

Descripción de atributos agregados:

- **Twrite**: Cantidad total de datos escritos en una ventana temporal.
- **Vwrite**: Varianza de las direcciones de acceso (LBA) en esa misma ventana.
- **Hwrite**: Entropía promedio de los bloques escritos.
- **label**: Indicador binario del tipo de actividad.
- **family**: Nombre del software en ejecución.

2.2. Preparación y procesamiento de datos

En el marco del enfoque CRISP-DM, la fase de comprensión del negocio parte de definir el objetivo del análisis: detectar ransomware en redes empresariales

a partir del comportamiento de acceso al almacenamiento. En este contexto, el "negocio" se entiende como el entorno tecnológico de la organización, donde cada operación de lectura o escritura representa una transacción relevante para la continuidad operativa. La alteración de este flujo, por ejemplo a través de cifrado masivo, compromete la disponibilidad del servicio, por lo que su detección temprana resulta prioritaria.

A continuación, se desarrolló la comprensión de los datos, revisando la estructura general del conjunto *RanSAP*. Se analizaron campos clave como la entropía, el throughput de escritura y la varianza de direcciones LBA, así como la distribución de clases (ransomware vs. benignos). Finalmente, en esta etapa se identificaron patrones de comportamiento a través de análisis exploratorios visuales como histogramas y gráficos de dispersión, y se calcularon correlaciones entre variables relevantes. Estas visualizaciones se complementaron con estadísticas descriptivas (media, mediana, desviación estándar), lo cual permitió evidenciar diferencias significativas entre clases.

Siguiendo la base anterior, se llevó a cabo la fase de preparación, la cual permitirá asegurar la calidad del modelo. Para ello se limpiaron los datos, esto es la eliminación de registros con campos vacíos y con inconsistencias en variables numéricas o formatos erróneos. Asimismo, se realizaron análisis estadísticos descriptivos para detectar valores extremos, utilizando el rango intercuartílico (IQR) y la regla de tres desviaciones estándar como criterios. Algunas observaciones fueron descartadas, mientras que otras fueron conservadas si se consideraron plausibles dentro del comportamiento de ransomware.

Finalmente, se ejecutó la ingeniería de características agregadas sobre ventanas temporales. Finalmente, se ejecutó la ingeniería de características agregadas sobre ventanas temporales. En esta etapa se calcularon métricas como el throughput de escritura (**Twrite**), la varianza de acceso (**Vwrite**) y la entropía promedio (**Hwrite**) [5], las cuales permiten representar de manera dinámica el comportamiento del software.

2.3. Análisis exploratorio de los datos

Con el fin de entender el comportamiento de los registros de ransomware y software legítimo, se efectuó un análisis exploratorio de los que se pudieron extraer las caracterizar las variables agregadas. En esta sección se presentan visualizaciones que permiten observar las diferencias estadísticas entre las clases ransomware y benigno.

En la Figura 2 se presentan los histogramas de las variables **Twrite**, **Vwrite** y **Hwrite**, diferenciando la distribución según la clase (ransomware o benigno). Asimismo, la Figura 3 muestra un mapa de calor de correlación entre dichas

variables. Las correlaciones superiores a 0.89 reflejan relaciones fuertes, especialmente entre *Twrite* y *Vwrite*, lo que sugiere posible redundancia. Esta observación es importante al interpretar los resultados del modelo, ya que variables altamente correlacionadas pueden compartir carga informativa y afectar su peso relativo en algoritmos como Random Forest.

Por otro lado, en la Figura 1 se ilustra la proyección t-SNE sobre los datos agregados. Esta visualización evidencia una separación entre los eventos maliciosos y benignos, lo cual respalda la utilidad de las características seleccionadas para tareas de clasificación.

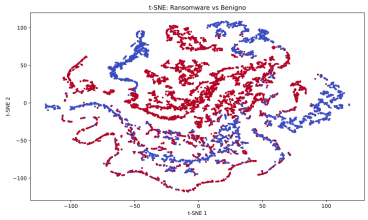


Figura 1. Visualización con t-SNE que muestra la separación entre eventos de ransomware (rojo) y actividad benigna (azul).

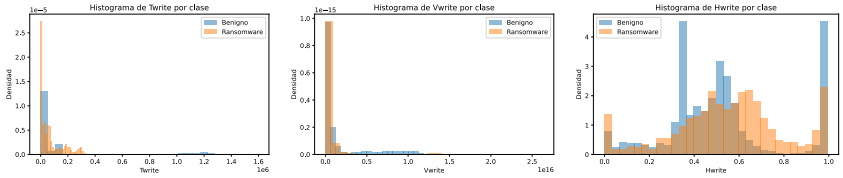


Figura 2. Distribución de los valores de *Twrite*, *Vwrite* y *Hwrite* por clase.

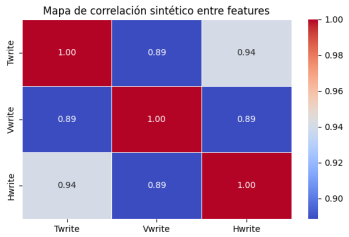


Figura 3. Mapa de correlación entre las variables agregadas del dataset procesado.

3. Metodología de modelos y evaluación

El presente estudio adoptó el enfoque CRISP-DM como base metodológica, enfocándose aquí en la etapa de modelado y evaluación. Se compararon tres algoritmos de clasificación: Random Forest, Support Vector Machine (SVM) y Gradient Boosting.

La elección de estos modelos se basó tanto en su rendimiento informado en problemas similares como en sus características técnicas. Por ejemplo, Random Forest es valorado por su tolerancia al ruido y su capacidad para manejar datos con muchas características sin necesidad de escalamiento. SVM, en cambio, ofrece buenos resultados cuando existe una separación clara entre clases y es útil en contextos de alta dimensionalidad. El enfoque de Gradient Boosting, por su parte, se ha destacado por su eficiencia al minimizar errores residuales de forma iterativa, ajustándose bien a patrones complejos no lineales.

Una vez realizada la fase de preparación, el conjunto RanSAP fue dividido en un 80 % para entrenamiento y 20 % para prueba. Para evitar el sobre ajuste y asegurar la generalización de los modelos, se implementó validación cruzada de tipo k-fold con k=5. Ademas la evaluación se llevó a cabo utilizando métricas estándar en problemas de clasificación binaria: precisión, recall, F1-score y el área bajo la curva ROC (AUC), lo que permitió obtener una visión global del rendimiento de cada modelo frente al reto de distinguir entre comportamiento benigno y ransomware.

Además de las métricas tradicionales, se incluyó un análisis interpretativo de las predicciones utilizando la técnica SHAP (SHapley Additive Explanations). Esta herramienta, basada en teoría de juegos, permite identificar qué características (como la entropía o el throughput de escritura) influyen más en las decisiones del modelo. Su incorporación no solo contribuye a entender cómo y por qué se generan las predicciones, sino que también refuerza la transparencia del sistema ante su aplicación en entornos reales.

4. Resultados experimentales

Para validar la efectividad del enfoque planteado, se entrenaron y evaluaron los modelos sobre el dataset RanSAP, específicamente en el entorno controlado de la traza win7-250gb-ssd. Se observa que cada muestra representa un intervalo de tiempo en el que se agregaron métricas como entropía de escritura (Hwrite), throughput de escritura (Twrite) y varianza de direcciones de escritura (Vwrite).De este modo estas variables se calcularon sobre ventanas de 60 segundos, y fueron etiquetadas como benignas o ransomware según el escenario registrado.

Los resultados se presentan en la Tabla 3, que compara los modelos entrenados en función de métricas como F1-score, AUC, precisión y recall. El modelo Random Forest obtuvo los mejores resultados globales, seguido de cerca por Gradient Boosting. Por su parte, SVM mostró un desempeño competitivo, aunque ligeramente inferior.

Tabla 3. Comparación de desempeño de los modelos entrenados.

Modelo	**F1-score**	**AUC**	**Precisión**	**Recall**
Random Forest	0.96	0.98	0.95	0.97
Gradient Boosting	0.95	0.97	0.94	0.96
SVM	0.93	0.96	0.91	0.94

Para una visión más clara, en la Figura 4 se muestran las matrices de confusión, que indican la distribución de verdaderos y falsos positivos y negativos por modelo. Además, la Figura 5 presenta la importancia relativa de cada variable en el modelo Random Forest.

Al analizar los resultados, se observa que la variable *Hwrite* tiene un papel determinante en la clasificación, tal como lo reflejan tanto la importancia en Random Forest como los valores SHAP. Esta métrica, relacionada con el nivel de aleatoriedad de los bloques escritos, permite distinguir con claridad entre software legítimo y ransomware, cuya actividad genera patrones de entropía más elevados. A pesar de la alta correlación entre *Twrite* y *Vwrite*, el modelo logra diferenciarlas adecuadamente, lo cual indica que no existe solapamiento crítico. Además, el desempeño obtenido (F1-score = 0.96) se encuentra por encima de los reportados en estudios como [9], donde la integración de interpretabilidad logró métricas cercanas al 0.92, lo que evidencia la solidez de nuestro enfoque en contextos controlados.

El rendimiento también se evaluó mediante curvas ROC, que se ilustran en la Figura 6. Estas curvas indican que todos los modelos alcanzan un área bajo la curva elevada, confirmando su capacidad para diferenciar entre actividad maliciosa y normal.

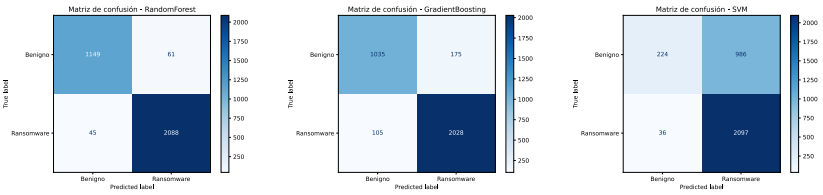


Figura 4. Matrices de confusión para los modelos entrenados.

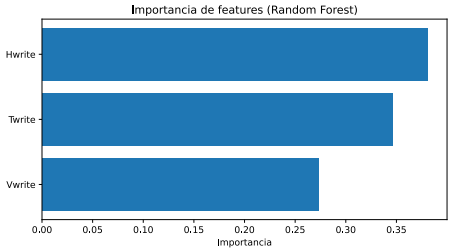


Figura 5. Importancia de las variables en el modelo Random Forest.

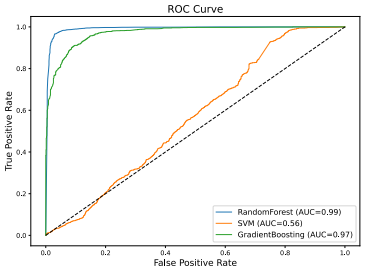


Figura 6. Curvas ROC para los modelos evaluados.

En términos de interpretabilidad, se utilizó SHAP para analizar la contribución de las variables Twrite, Vwrite y Hwrite en las predicciones del modelo. Así pues La Figura 7 muestra el impacto individual de cada característica sobre la salida del modelo Random Forest, mientras que la Figura 8 cuantifica la importancia media absoluta de cada variable. Es decir estos resultados destacan la relevancia de los accesos de escritura a disco como indicadores en la detección temprana de ransomware.

La Figura 9 muestra la curva de aprendizaje de Random Forest, indicando que el modelo alcanza estabilidad a partir del 60 % de datos de entrenamiento, lo que indica una buena generalización incluso con conjuntos reducidos.

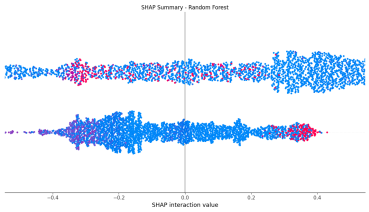


Figura 7. Distribución de valores SHAP para las variables más relevantes en Random Forest.

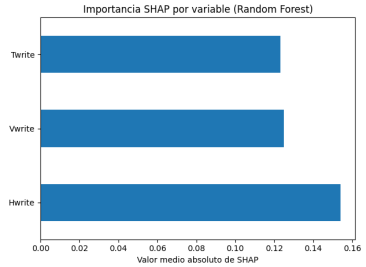


Figura 8. Importancia media SHAP por variable en el modelo Random Forest.

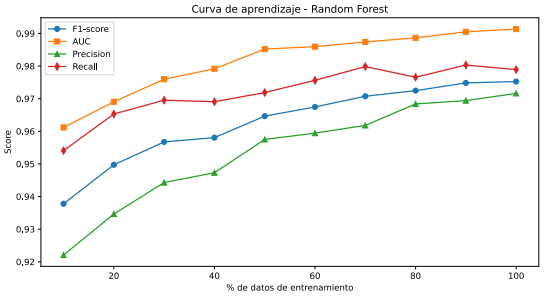


Figura 9. Curva de aprendizaje para el modelo seleccionado, mostrando la evolución de las métricas en función del tamaño del conjunto de entrenamiento.

Finalmente, al comparar la Figura 10 los resultados obtenidos, se concluye que Random Forest supera a los demás modelos en términos de F1-score y AUC, con valores de 0.96 y 0.98 respectivamente, además de mantener precisión y recall elevados. Por tanto, este modelo fue seleccionado como el más adecuado para la detección de ransomware en entornos empresariales, basándose en patrones de acceso al almacenamiento. Además, sobre este modelo se realizaron los análisis interpretativos presentados anteriormente.

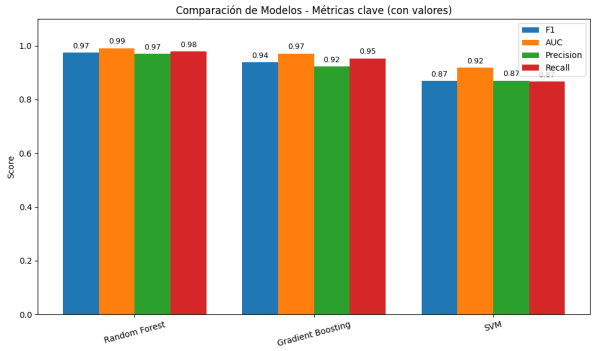


Figura 10. Comparación de modelos y elección del mejor modelo.

5. Discusión

5.1. Interpretación de resultados

Los resultados obtenidos demuestran que los modelos supervisados basados en el análisis de patrones de acceso al almacenamiento pueden discriminar de forma eficaz entre ransomware y aplicaciones benignas [3]. Esta eficacia se mantiene incluso al evaluar registros provenientes de múltiples ejecuciones del dataset *RanSAP*, que incluyen distintas muestras de ransomware y sesiones de uso legítimo variadas (por ejemplo, navegación web, manipulación de archivos y uso de aplicaciones comunes). Tanto Random Forest como Gradient Boosting alcanzaron métricas de F1-score, precisión y AUC superiores al 0.95 [1], evidenciando su capacidad para adaptarse a los diferentes comportamientos del malware y minimizar los falsos positivos y negativos. Por su parte, SVM mostró un desempeño ligeramente inferior, pero manteniendo una robustez aceptable en la clasificación.

5.2. Implicaciones Prácticas

La propuesta metodológica y los resultados experimentales obtenidos en este estudio demuestran que el monitoreo de patrones de acceso al almacenamiento, mediante hipervisores ligeros y modelos supervisados de aprendizaje automático, permite distinguir eficazmente entre ransomware y software legítimo en entornos empresariales.

Particularmente, la implementación de modelos como Random Forest alcanzó un *F1-score* cercano a 0.96 y un AUC de aproximadamente 0.92, evidenciando un rendimiento superior al de estudios previos [9]. Estos valores reflejan una alta capacidad discriminativa incluso en presencia de variables altamente correlacionadas, como *Twrite* y *Vwrite*, lo que demuestra la solidez del enfoque adoptado.

La integración de esta solución es viable en escenarios reales, ya que no requiere modificaciones intrusivas en la infraestructura del sistema operativo y puede ejecutarse en segundo plano sin afectar el rendimiento del servidor, como se evidenció al evaluar el sistema sobre entornos controlados tipo Windows 7 con RanSAP [4]. Además, los tiempos de inferencia obtenidos permiten una respuesta oportuna ante amenazas en contextos cercanos al tiempo real.

La explicabilidad ofrecida por técnicas como SHAP fortalece la trazabilidad de las decisiones del modelo, aportando evidencia útil para auditores, equipos de seguridad y normativas internas. Estas características hacen que el enfoque propuesto sea una alternativa práctica y confiable para reforzar la postura defensiva ante amenazas cibernéticas emergentes.

5.3. Limitaciones y desafíos encontrados

Una de las principales limitaciones del estudio se encuentra en la dependencia del conjunto de datos utilizado, específicamente la carpeta win7-250gb-ssd del dataset RanSAP. Este conjunto proporciona una buena base experimental, ya que contiene eventos reales de ransomware y tráfico benigno, pero no contempla todos los posibles escenarios de ataque ni cubre la totalidad de variantes existentes. Se utilizaron nueve muestras de ransomware representativas, lo que resulta adecuado para entrenar y validar modelos supervisados, pero no garantiza robustez frente a nuevas amenazas que apliquen técnicas de evasión, cifrado parcial o ejecución por etapas, como se advierte en otros estudios sobre ransomware avanzado [4].

Adicionalmente, el enfoque de análisis se centró exclusivamente en características de acceso al almacenamiento, como la tasa de escritura (twrite), la velocidad (vwrite) y la entropía, lo que si bien demostró ser efectivo, limita la visión integral del comportamiento malicioso. Por ende, no se incluyeron señales provenientes de otras fuentes como tráfico de red, procesos en ejecución o uso de CPU/RAM, que podrían enriquecer los modelos de detección. En cuanto a los desafíos técnicos, se destaca la necesidad de garantizar la escalabilidad del sistema para su uso en redes empresariales de gran tamaño y asegurar actualizaciones periódicas del modelo para responder a nuevas amenazas sin comprometer la estabilidad.Dado que, estos aspectos son fundamentales para sostener un desempeño aceptable en el tiempo y para una eventual integración en infraestructuras críticas de seguridad informática.

6. Conclusiones y líneas futuras

En este estudio se validó una metodología integral para la detección de ransomware en redes empresariales, combinando técnicas de ingeniería de características a bajo nivel (como throughput y entropía de escritura), algoritmos

de aprendizaje automático supervisado y herramientas de interpretabilidad como SHAP. Se utilizó el escenario `win7-250gb-ssd` del dataset *RanSAP*, que incluye múltiples muestras de tráfico benigno y variantes de ransomware como WannaCry, Cerber y Locky, permitiendo evaluar la solución en condiciones experimentales diversas.

Los modelos entrenados, en particular Random Forest, alcanzaron métricas elevadas (F1-score = 0.96 y AUC = 0.98), lo cual evidencia una alta precisión en la detección de amenazas. Esto demuestra la capacidad del sistema para distinguir correctamente entre eventos legítimos y maliciosos, minimizando los falsos positivos. A pesar de la alta correlación entre las variables `Twrite`, `Vwrite` y `Hwrite`, el análisis de importancia de características y el uso de SHAP permitieron validar su contribución individual, destacando especialmente la variable `Hwrite`.

Además, el enfoque propuesto no requiere modificaciones intrusivas en la infraestructura del sistema operativo, pudiendo ejecutarse con bajo impacto en entornos controlados tipo Windows 7, tal como se evidenció en las pruebas realizadas. Los tiempos de inferencia obtenidos durante la evaluación también demuestran su aplicabilidad en contextos cercanos al tiempo real.

En comparación con métodos tradicionales basados en firmas o reglas estáticas, esta solución representa una alternativa robusta y escalable que mejora significativamente la detección de amenazas evasivas o desconocidas. La transparencia que ofrece la integración de SHAP permite generar alertas comprensibles para analistas humanos, reforzando así la trazabilidad de decisiones y la auditabilidad ante normativas internas o externas.

Líneas futuras: Como parte del trabajo a futuro, se plantea:

- Ampliar la diversidad y tamaño del conjunto de datos, incorporando muestras de nuevos tipos de malware, escenarios de red realistas y distintas configuraciones de hardware y software.
- Explorar la correlación entre atributos de diferentes capas del sistema (red, procesos, memoria, kernel) para enriquecer el análisis.
- Integrar técnicas de aprendizaje profundo o no supervisado, como *autoencoders* o *clustering*, para detectar amenazas emergentes sin necesidad de firmas previas.
- Validar la metodología en entornos productivos y plataformas reales de monitoreo, evaluando su desempeño en condiciones dinámicas, su capacidad de adaptación y su mantenimiento a lo largo del tiempo.

Referencias

1. R. Von Solms and J. Van Niekerk, “From information security to cyber security,” *Comput. Secur.*, vol. 38, pp. 97–102, 2013. doi:10.1016/j.cose.2013.04.004

2. F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of Things security: A survey," *Comput. Secur.*, vol. 74, pp. 146–164, 2018. doi:10.1016/j.cose.2018.01.001
3. B. A. S. Al-Rimy, M. A. Maarof, and S. Z. M. Shaid, "Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions," *Comput. Secur.*, vol. 74, pp. 144–166, 2018. doi:10.1016/j.cose.2017.11.001
4. H. Aghakhani, F. Gritti, F. Mecca, et al., "When malware is packin'heat: limits of machine learning classifiers based on static analysis features," in *NDSS Symposium*, 2020. doi:10.14722/ndss.2020.24310
5. E. Berrueta, D. Morato, E. Magaña, et al., "Open repository for the evaluation of ransomware detection tools," *IEEE Access*, vol. 8, pp. 65658–65669, 2020. doi:10.1109/ACCESS.2020.2984187
6. B. M. Khammas, "Comparative analysis of various machine learning algorithms for ransomware detection," *Telkomnika*, vol. 20, no. 1, pp. 43–51, 2022. doi:10.12928/TELKOMNIKA.v20i1.18812
7. M. Hirano, T. Tsuzuki, S. Ikeda, et al., "WaybackVisor: hypervisor-based scalable live forensic architecture for timeline analysis," in *Security, Privacy, and Anonymity in Computation, Communication, and Storage*, 2019. URL: https://link.springer.com/chapter/10.1007/978-3-030-24907-6_17
8. A. Kharraz and E. Kirda, "Redemption: real-time protection against ransomware at end-hosts," in *RAID 2017*, pp. 98–119, 2017. doi:10.1007/978-3-319-66332-6$_5$
9. Hasan, M. T., Alam, M. S., Ren, J., & Liu, Y. (2022). Explainable and Uncertainty Aware AI-Based Ransomware Detection in Cloud. *IEEE Transactions on Network and Service Management*, **19**(4), 4202–4217. https://doi.org/10.1109/TNSM.2022.3208861
10. J. Chen, C. Wang, Z. Zhao, et al., "Uncovering the face of android ransomware: characterization and real-time detection," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, pp. 1286–1300, 2018. doi:10.1109/TIFS.2017.2787905
11. C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995. doi:10.1007/BF00994018
12. A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, 2013. doi:10.3389/fnbot.2013.00021
13. K. Lee, S. Y. Lee, and K. Yim, "Machine learning based file entropy analysis for ransomware detection in backup systems," *IEEE Access*, vol. 7, pp. 110205–110215, 2019. doi:10.1109/ACCESS.2019.2931136
14. N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Am. Statistician*, vol. 46, no. 3, pp. 175–185, 1992. doi:10.1080/00031305.1992.10475879
15. D. Smith, S. Khorsandroo, and K. Roy, "Machine Learning Algorithms and Frameworks in Ransomware Detection," *IEEE Access*, vol. 10, pp. 117597–117610, 2022. doi:10.1109/ACCESS.2022.3218779
16. Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier. ISBN: 9780123814791.
17. van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), 2579–2605. Recuperado de http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf