

TÉCNICAS DE MACHINE LEARNING PARA LA DETECCIÓN DE RANSOMWARE: REVISIÓN SISTEMÁTICA DE LITERATURA

SANDRA HARO



INTRODUCCIÓN



El objetivo principal de la presente investigación fue identificar los algoritmos del Machine Learning para la detección y clasificación de las diferentes familias ransomware, así como las herramientas de software a utilizar para el entrenamiento del conjunto de datos con las cuales se puede identificar patrones, logrando así, brindar un apoyo en la detección de ransomware y de esta manera mitigar daños irreparables a la integridad de la información.

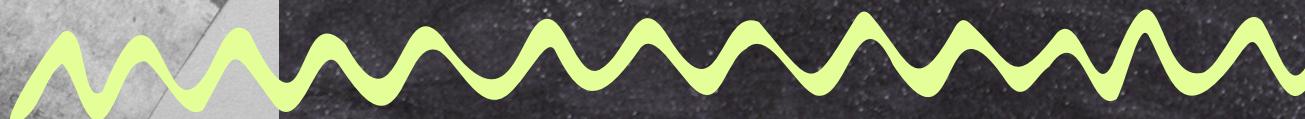


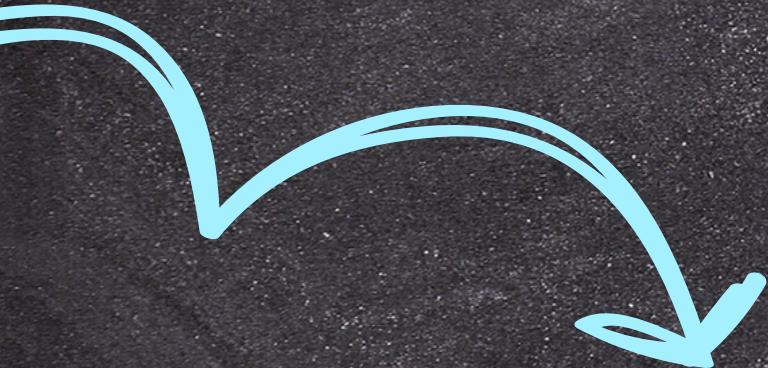


METODOLOGÍA

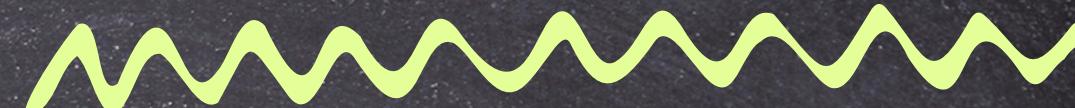
Para la planificación y conducción de la Revisión Sistemática de Literatura (RSL), se aplicó la metodología de Barbara Kitchingham cuyas fases y tareas se describen a continuación:

- Planificar la revisión sistemática de literatura con el uso de la herramienta Parsifal.
- Desarrollar la revisión sistemática de literatura con la planificación definida.
- Documentar e interpretar los resultados de la revisión





DISCUSIÓN Y RESULTADOS



Planificar la revisión sistemática de literatura.

A continuación, se detallan cada una de las tareas realizadas durante la etapa de planificación de a revisión sistemática de literatura:

Determinar las preguntas de investigación



Preguntas del Mapeo Sistemático (MQ).



MQ1. ¿Cuántos estudios se han publicado en los últimos cinco años acerca de las técnicas de machine learning para la detección de ransomware?

MQ2. ¿Cuáles son los autores más relevantes y activos en este ámbito de estudio?

MQ3. ¿En dónde se han publicado la mayor cantidad de los artículos sobre el tema de estudio?



Preguntas de la Revisión Sistémica (RQ).



RQ1. ¿Cuáles son los algoritmos del machine learning que mejores resultados obtuvieron para la detección de ataques ransomware?

RQ2. ¿Cómo ayuda el machine learning en la prevención de ataques ransomware?

RQ3. ¿Qué herramientas de software se han utilizado para la aplicación de técnicas y/o algoritmos de machine learning en la detección de ransomware?

Establecer proceso de búsqueda

Se definieron los términos base aplicando el método PICOC para definir el ámbito de la SLR. Sus componentes son:

Población (P): "Ransomware"

Intervención (I): "Machine Learning" OR "Deep Learning"

Comparación (C): No aplica

Resultados (O): "Algorithms" OR "Methods" OR "Techniques" OR "Classification"

Contexto (C): "Intelligence Artificial"

A continuación se detalla la lista de las palabras claves sinónimas..

Tabla 1: Definición de palabras clave

Palabra Clave	Sinónimos	Relación PICOC
Algorithms	Classification Detection Methods Techniques	Resultados
Machine Learning	Deep Learning	Intervención
Ransomware	s/n	Población

Cadenas de Búsqueda

Se consideró palabras claves a partir de una revisión preliminar de artículos, mediante el uso de Thesaurus de IEEE se buscó sinónimos, se utilizó los operadores lógicos "AND/OR" con la finalidad de generar las cadenas de búsqueda. La herramienta Parsifal generó la cadena de búsqueda general, la misma que se modificó con cada base de datos y todas las palabras claves se definieron en idioma inglés, mismas que se encuentran estructuradas de la siguiente manera

Tabla 3. Cadenas de búsqueda

Base de datos	Cadenas de Búsqueda
Parsifal	("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")
ACM	((("Ransomware") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")) Publication Date: (01/01/2017 TO 12/31/2021)
IEEE	((("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" OR "Classification" OR "Detection" OR "Methods" OR "Techniques")) Filters Applied: 2017-2021

Science@Direct	((("Ransomware") AND ("Machine Learning" AND "Deep Learning") AND ("Algorithms" AND "Classification" OR "Detection" AND ("Methods" OR "Techniques"))) Year: 2017-2021 Type: Research articles
Scopus	TITLE-ABS-KEY(("Ransomware") AND ("Machine Learning" OR "Deep Learning") AND ("Algorithms" AND ("Classification" OR "Detection") AND ("Methods" OR "Techniques")))) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"cp")) AND (LIMIT-TO (SUBJAREA,"COMP")) AND (LIMIT-TO (PUBYEAR,2021) OR LIMIT-TO (PUBYEAR,2020) OR LIMIT-TO (PUBYEAR,2019) OR LIMIT-TO (PUBYEAR,2018) OR LIMIT-TO (PUBYEAR,2017))

Verificación de calidad

Las siguientes preguntas se establecieron para evaluar la calidad de los artículos preseleccionados:

QA1. ¿El autor cita o utiliza alguna técnica y/o algoritmo del machine learning para la detección de ataques ransomware?

QA2. ¿En los estudios se menciona alguna herramienta de software utilizada para la aplicación de técnicas y/o algoritmos para la detección de ransomware?

QA3. ¿En los artículos se habla acerca de la importancia del machine learning para la detección de ataques ransomware?

Para finalizar se determinaron los parámetros de puntuación para definir que artículos serán seleccionados y rechazados. Los parámetros se detallan a continuación:

- Si la respuesta es Si su puntuación será 1.0
- Si la respuesta es Parcialmente su puntuación será 0.5
- Si la respuesta es No su puntuación será 0.0

Evaluación de calidad a los artículos primarios

Los 383 artículos obtenidos fueron revisados y analizados en su título y resumen, tomando en consideración los criterios de inclusión y exclusión. Del total se eliminaron 312 artículos que son irrelevantes al objeto de estudio, además se descartaron porque su argumentación referente a técnicas de machine learning para la detección de ransomware es débil y no dan contestación a las preguntas de investigación planteadas. Obteniendo 71 artículos para su revisión y análisis.

Se aplicaron las preguntas de calidad a cada uno de los 71 artículos seleccionados, , y se seleccionaron 34 los cuales tenían una calificación igual o superior a 2 puntos

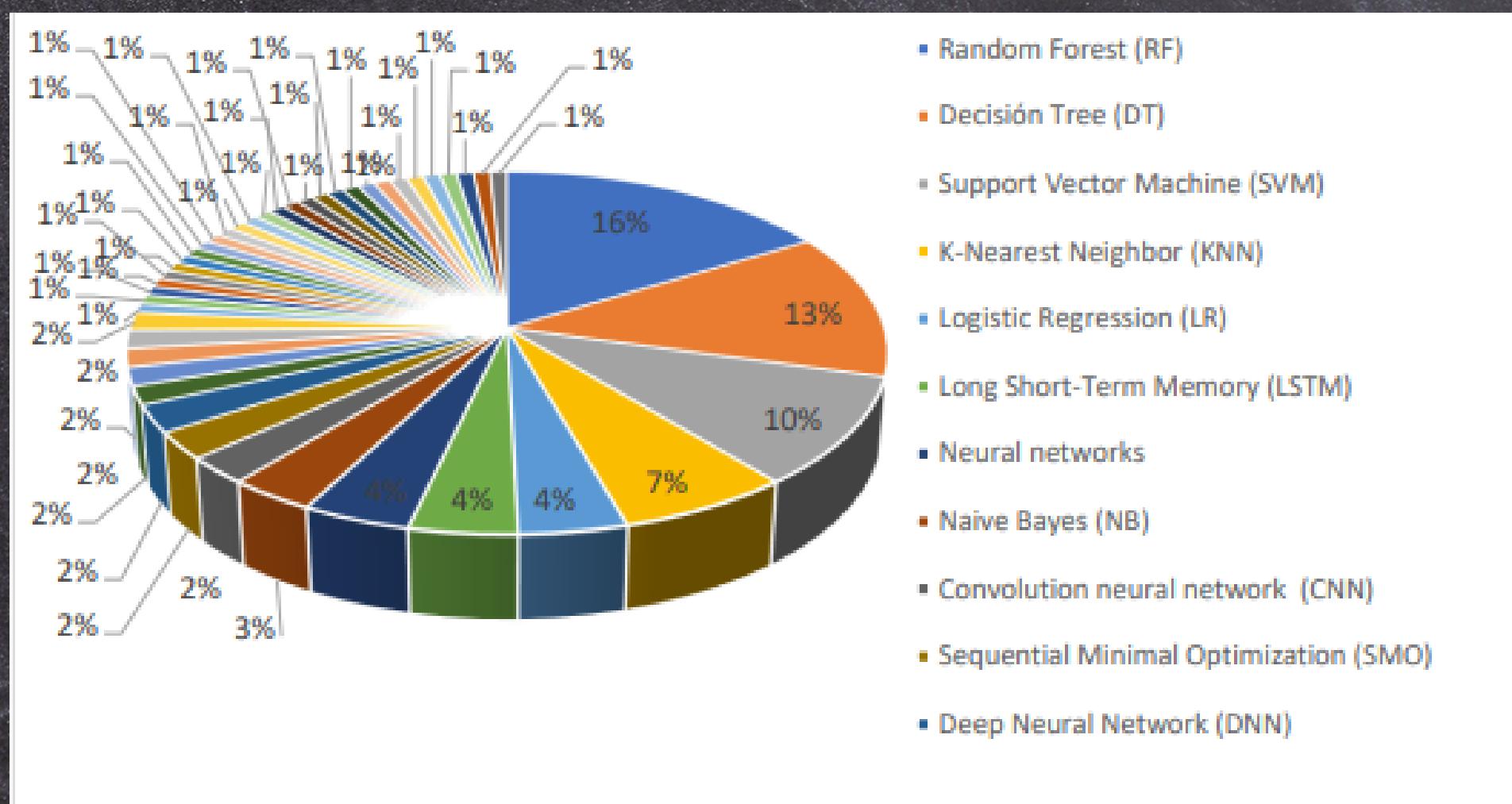
Tabla 4. Artículos seleccionados

Código	Título	Referencia
SA01	A Novel Malware Analysis for Malwarem Detection and Classification Using Machine Learning Algorithms	(Sethi et al., 2017)
SA02	A Survey on Malware Detection with Deep Learning.	(Sahin y Bahtiyar, 2020)
SA03	Analysis of Machine Learning Techniques for Ransomware Detection.	(Noorbehbahani et al., 2019)
SA04	An Empirical Comparison of Supervised Algorithms for Ransomware Identification on Network Traffic.	(Manzano et al., 2020)
SA05	Attention in Recurrent Neural Networks for Ransomware Detection.	(Agrawal et al., 2019)
SA06	Android Ransomware Detection using Machine Learning Techniques: A Comparative Analysis on GPU and CPU.	(S. Sharma et al., 2020)
SA07	A Digital DNA Sequencing Engine for Ransomware Detection Using Machine	(Khan et al., 2020)



ANÁLISIS Y CLASIFICACIÓN DE LA INFORMACIÓN

Análisis de algoritmos: Se tomó en cuenta el número total de artículos donde se menciona los algoritmos utilizados, estos datos se los representó en un grado porcentual para hacer visible de una mejor manera la información



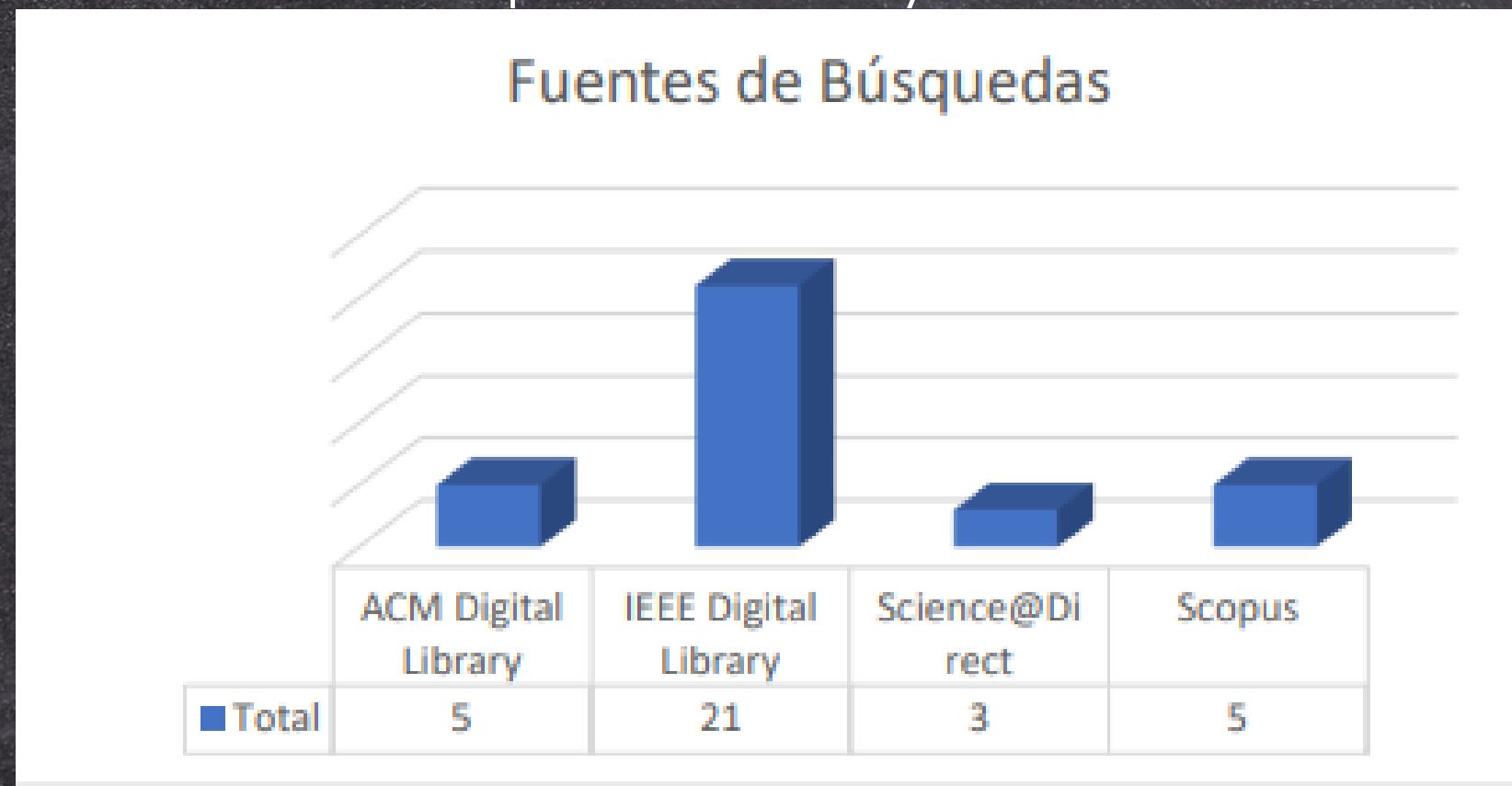


INFORME DEL MAPEO SISTEMÁTICO

En el siguiente apartado se dan a conocer las respuestas de las preguntas de mapeo sistemático (MQ).

MQ1. ¿Cuántos estudios se han publicado en los últimos cinco años acerca de las técnicas de machine learning para la detección de ransomware?

En cuanto a MQ1, tomando en cuenta los últimos cinco años se han publicado 34 artículos, siendo IEEE Digital Library el mayor aportador de publicaciones científicas con 21 aportaciones, seguido de ACM Digital Library y Scopus con 5 aportaciones cada uno respectivamente y Science direct con 3 aportaciones.

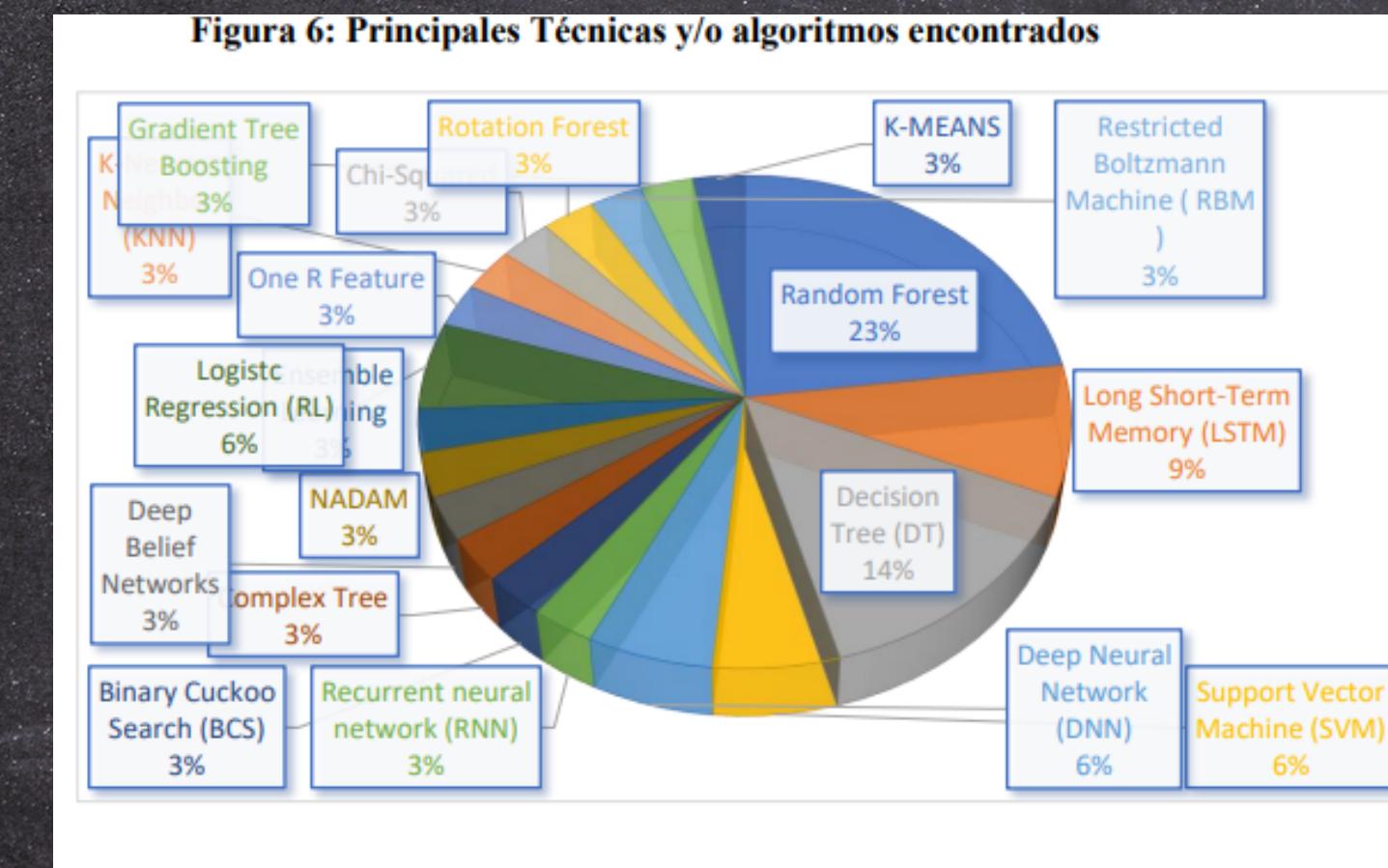




Análisis de la revisión sistemática

RQ1. ¿Cuáles son los algoritmos del machine learning que mejores resultados obtuvieron para la detección de ataques ransomware?

Para poder dar contestacion a RQ1, se revisó y analizó los 34 artículos seleccionados, donde se identificó 19 algoritmos y/o técnicas de machine learning utilizados para la detección de ataques ransomware. De cada artículo se extrajo el algoritmo que mejor rendimiento aporto al investigador.

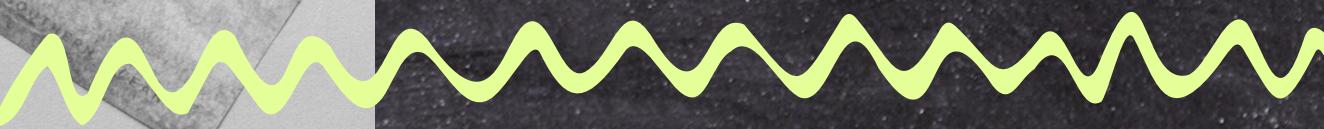




CONCLUSION

EL USO DE LA HERRAMIENTA PARSIFAL PERMITIÓ EL DESARROLLO ÓPTIMO Y EFICIENTE DEL PROCESO DE LA PLANIFICACIÓN DE LA REVISIÓN SISTEMÁTICA DE LITERATURA, LO QUE AYUDÓ A GESTIONAR LA INVESTIGACIÓN DE MANERA ORDENADA, SIGUIENDO UNA SERIE DE PASOS PRECISOS Y ESPECÍFICOS PARA EXTRAER LA INFORMACIÓN DE LOS ARTÍCULOS RELEVANTES QUE CUMPLIERON CON LOS CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN DEFINIDOS CON EL FIN DE IDENTIFICAR LAS TÉCNICAS DEL MACHINE LEARNING PARA LA DETECCIÓN DE RANSOMWARE.

DE LOS 34 ARTÍCULOS QUE FUERON ANALIZADOS SE OBTUVO QUE LAS TÉCNICAS MÁS UTILIZADAS PARA LA DETECCIÓN DE RANSOMWARE SON: EL ALGORITMO RANDOM FOREST (RF) CITADO EN LA MAYOR PARTE DE ARTÍCULOS SELECCIONADOS COMO EL QUE MEJOR RENDIMIENTO OBTIENE EN LA DETECCIÓN Y CLASIFICACIÓN DE RANSOMWARE, SEGUIDO DE DECISION TREE (DT) Y LONG SHORT-TERM MEMORY (LSTM).



iGRACIAS!