UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS

AUTOR:

Icochea Rivera, George Eduard (orcid.org/0000-0003-0400-4044)

ASESOR:

Dr. Alfaro Paredes, Emigdio Antonio (orcid.org/0000-0002-0309-9195)

LÍNEA DE INVESTIGACIÓN:

SISTEMA DE INFORMACIÓN Y COMUNICACIONES

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo sostenible y adaptación al cambio climático

LIMA – PERÚ

2022

Dedicatoria

Dedico esta investigación a mis tías Sonia Rivera y Lucia Rivera, quienes fueron mi soporte e inspiración para seguir adelante. Su apoyo y dedicación fue un valioso pilar en mi formación y me ayudó a superar los obstáculos que se presentaban a lo largo de la carrera.

Agradecimiento

Agradezco a Dios por las circunstancias favorables que me brindo, también agradezco a mi papa, hermanas y tíos que me apoyaron de forma incondicional. Del mismo modo agradezco a mi asesor Dr. Emigdio Alfaro y a los profesores que me acompañaron a lo largo de la carrera profesional con sus enseñanzas y guía.

Índice de contenidos

Índice de tablas

Índice de figuras

Índice de anexos

Índice de abreviaturas

Sigla	Significado en ingles	Significado en español	Pág.
CNN	convolutional neural network	red neuronal convolucional (Al-Alyan et al., 2020, p. 2753)	9
URL	Uniform Resource Locator	localizador de recursos uniforme (Al-Alyan et al., 2020, p. 2753)	2
RAM	Random Access Memory	Memoria de Acceso Aleatorio (Flygare y Holmqvist, 2017, p. 12)	4
CPU	Central Processing Unit	Unidad Central de Procesamiento (Verona et al., 2016, p. 281)	4
RPA	Robotic Process Automation	Automatización robótica de procesos (PremaLatha et al., 2020, p. 4876)	2
SVM	Support Vector Machines	Máquina de vector de soporte (Céspedes, 2021, p. 31)	10

Resumen

Esta investigación se desarrolló con la finalidad de realizar una aplicación de Automatización robótica de procesos para la clasificación de correos electrónicos como phishing y noticias falsas basada en el algoritmo fusionado ELM - K-step CD-K. Se seleccionó investigaciones de diversas revistas y libros con una perspectiva de elaboración, técnicas y métodos de clasificación de correos. EL objetivo es que tener los mejores resultados en los indicadores como tiempo de clasificación, eficacia, uso de CPU y de memoria RAM. Se empleó un RPA con cuatro módulos: (a) actualización de la base de datos de URL maliciosos y que contengan noticias falsas, (b) análisis de los correos que contenga una URL y descarta a los otros, (c) validación de los correos que tenga URL no sean maliciosos y (d) traslado de todos los correos con URL maliciosos a una carpeta predefinida.

Luego se presentó los resultados de uso del RPA. Posteriormente se presentó una comparación. El RPA con el algoritmo ELM - K-step CD-K obtuvo 0.34 en segundos para el tiempo de clasificación. Del mismo modo, se obtuvo un porcentaje de eficacia de 100%. Asimismo, el porcentaje de uso del CPU fue de 29.57%, mientras que en el modelo de SVM fue de 17.64%. De igual manera, el uso de memoria RAM fue de 286.36 MB, mientras que en el modelo de XGBoost fue de 4.00 MB. Por último, se presentó las recomendaciones para futuras investigaciones como evaluar más indicadores, centrar la investigación en redes sociales, comparar otras herramientas de automatización, tener una implementación web y utilizar otros sistemas de aprendizaje profundo para la mejora en la clasificación.

Palabras clave: RPA, URL, Automatización robótica de procesos, phishing, noticias falsas, correos electrónicos, Automatización

Abstract

This research was developed with the purpose of carrying out a Robotic Process Automation application for the classification of emails as phishing and fake news based on the merged algorithm ELM - K-step CD-K. Research from various magazines and books was selected with a perspective of elaboration, techniques and methods of mail classification. The goal is to have the best results in indicators such as classification time, efficiency, CPU and RAM usage. An RPA with four modules was used: (a) update of the database of malicious URLs and those containing false news, (b) analysis of the emails that contain a URL and discard the others, (c) validation of the emails with non-malicious URLs and (d) moving all emails with malicious URLs to a predefined folder.

Then the results of using the RPA were presented. A comparison was later presented. The RPA with the ELM - K-step CD-K algorithm obtained 0.34 seconds for the classification time. In the same way, an efficiency percentage of 100% was obtained. Also, the percentage of CPU usage was 29.57%, while in the SVM model it was 17.64%. Similarly, the use of RAM memory was 286.36 MB, while in the XGBoost model it was 4.00 MB. Finally, recommendations for future research were presented, such as evaluating more indicators, focusing research on social networks, comparing other automation tools, having a web implementation, and using other deep learning systems to improve classification.

Keywords: RPA, URL, Robotic Process Automation, Phishing, Fake News, Emails, Automation.

I. INTRODUCCIÓN

En este capítulo se precisa los estudios sobre la aplicación de automatización robótica de procesos para la clasificación de correos electrónicos y priorizando la separación del phishing como también la separación de las fake news. Se describirá los puntos que no han sido desarrollados en los trabajos anteriores y que se está abarcando con la nueva investigación. Asimismo, la importancia y lo que conlleva no tener el RPA de clasificación de correos como phishing y fake news.

Para el respectivo sustento de la aplicación RPA para la clasificación de correos electrónicos encontramos los siguientes temas: (a) automatización del cliente de correo electrónico con RPA (Akshay et al., 2020) y (b) automatización de correo electrónico mediante automatización robótica de procesos (PremaLatha et al., 2020). Asimismo, para el tema de phishing se encontró lo siguiente: (a) la detección robusta de suplantación de identidad de URL basada en aprendizaje profundo (Al-Alyan et al., 2020) y (b) un esquema de clasificación de entropía máxima para la detección de phishing utilizando características parsimoniosas (Asani et al., 2021). Para el tema de fake news se encontró la automatización del proceso de detección de noticias falsas (Nayak et al., 2019).

En este caso, no se ha localizado estudios con una correcta clasificación de correo en Phishing utilizando automatización robótica de procesos. Tampoco se ha localizado investigaciones que combinen una clasificación de correos utilizando automatización robótica de procesos entre Phishing y Fake News.

Teniendo en cuenta los estudios ya efectuados y que no se ha encontrado una aplicación de automatización robótica de procesos con una buena clasificación de correos que contengan Phishing y Fake News juntos, se está planteando esta investigación con dicha combinación con la finalidad de determinar el efecto combinado de estas clasificaciones con un RPA.

La carencia de estudios sobre esta aplicación de automatización robótica de procesos para la clasificación de correos de Phishing y Fake News ha contribuido a que la detección y la defensa contra el phishing web sea una tarea de investigación urgente y esencial (Yi et al., 2018). Asimismo, es más fácil distribuir noticias falsas a través de las redes sociales, que a través de los medios tradicionales que tienden a verificar y validar las noticias antes de su distribución (Nayak et al., 2019).

Las justificaciones que se tuvieron en cuenta para este estudio fueron: teórica, económica y social. Se tiene una justificación teórica para un mejor aporte del conocimiento en esta investigación. Se tiene una justificación económica ya que permitirá reducir el tiempo y por consecuencia los costos empleados para dicha función. Por último, se tiene una justificación social porque aporta a la disminución de contenido engañoso para los usuarios de correo, como lo es el phishing y las fake news.

La justificación teórica tiene como base el aporte de conocimientos. PremaLatha et al. (2020) indicaron que el objetivo principal de la automatización robótica de procesos (RPA) es realizar tareas repetitivas sin interacción humana, ahorrando tiempo y costo, empoderamiento de los empleados y se puede producir los resultados precisos sin ninguna desviación en términos de calidad, brindando el resultado perfecto (p. 4876).

En relación a la justificación económica, Akshay et al. (2020) comentaron que dada la carga de trabajo que generan los correos electrónicos, desde hace mucho tiempo se desea automatizar varios aspectos del procesamiento del correo electrónico (p. 787). Si una empresa u organización de soporte requiere responder un centenar de correos electrónicos que son recibidos de forma masiva en sus bandejas de entrada para obtener una respuesta, entonces esa organización podría escapar de tal situación y hacer que RPA se encargue de ellos (Akshay et al., 2020, p. 787). Los correos electrónicos se pueden segmentar en grupos y la solución RPA puede responder a estos correos electrónicos y por lo tanto, los correos que no pueden ser segmentados en un grupo pueden ser atendidos por el personal respectivo (Akshay et al., 2020, p.787).

Esta investigación se justificará socialmente a fin de reducir el contenido engañoso. Mertoğlu et al. (2020) comentaron que más importante es el contenido potencial de noticias falsas entregado por estas fuentes; que pueden engañar a la sociedad y causar disturbios sociales como desencadenar la violencia contra las minorías étnicas y los refugiados, causar temores innecesarios relacionados con problemas de salud o incluso, a veces, resultar en crisis, disturbios devastadores y huelgas (Mertoğlu et al., 2020, p.1).

En base a la realidad problemática presentada en esta investigación se planteó el problema general y los problemas específicos de la investigación.

Como problema general se tuvo: ¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K? Los problemas específicos fueron los siguientes:

- PE1: ¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el tiempo de clasificación?
- PE2: ¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en la eficacia de la clasificación?
- PE3: ¿Cuál fue el resultado del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el Uso de CPU?
- PE4: ¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el Uso de memoria RAM?

Por eso, el objetivo general fue determinar el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K. En esta investigación, los objetivos específicos propuestos se mencionarán a continuación:

- OE1: Determinar el resultado del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el tiempo de clasificación.
- OE2: Establecer cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en la eficacia de la clasificación.
- OE3: Estimar cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el uso de CPU.

- OE4: Medir cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el uso de RAM.

Se espera validar el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K (Akshay et al., 2020, p. 794; PremaLatha et al., 2020, p. 4882).

- HE1: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el tiempo de clasificación.
  Akshay et al. (2020) indicaron que este enfoque redujo el tiempo de clasificación; esto permitió a los usuarios tener todo el control sobre las tareas asignadas (p. 788). PremaLatha et al. (2020) comentaron que el tiempo se redujo y mejoro la precisión, dejando tiempo disponible para los representantes y evitando las tareas repetitivas y aburridas (p. 4876).

- HE2: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K incremento la eficacia de la clasificación.
  Akshay et al. (2020) propusieron un modelo de clasificación con la que revisan todos los correos nuevos que ingresan a su bandeja de entrada, los segmenta e identificar correos electrónicos no deseados (p. 793). Prexawanprasut et al. (2017) también recalcaron que implementaron herramientas de administración de correo electrónico que clasifican los mensajes como información útil (p. 784).

- HE3: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de CPU.
  Prexawanprasut et al. (2017) indicaron que su modelo podra economizar recursos a expensas de potencia de CPU, reduciendo el uso de dicho recurso (Prexawanprasut et al., 2017, p. 785). La medición del uso de

CPU es antes y durante la ejecuacion de cada modelo a comparar (Soto et al., 2020, p.36).

- HE4: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de RAM.
  La medición del uso de memoria RAM es antes y durante la ejecución de cada modelo a comparar (Soto et al., 2020, p.36). Tambien Verona et al. (2016) indicaron el uso que tiene la memoria RAM tiene un impacto en el rendimiento de la aplicación a implantar (p. 281).

II. MARCO TEÓRICO

En el presente capitulo se mostrará los antecedentes de estudios similares resaltando los objetivos y conclusiones, también se presentará las teorías relacionadas que incluirán concepto variables y dimensiones. Por último, los marcos conceptuales en relación a la investigación abarcada.

A continuación, indicaremos los antecedentes de relacionado a la investigación. Entre ellas se tiene a PremaLatha et al. (2020) que publicaron *Email Automation Using Robotic Process Automation*. De igual forma tenernos a Akshay et al. (2020) con su publicación de *email client automation with RPA*. Borg (2021) con "*Clasificación de correo electrónico con aprendizaje automático e incrustaciones de palabras para mejorar la atención al cliente"*, También tenemos a Al-Alyan et al. (2020) con "*Detección robusta de phishing de URL basada en aprendizaje profundo"* y a Dutta (2021) con "*Detecting phishing websites using machine learning technique"*. Por ultimo a Mertoğlu et al. (2020) con su publicación "*Detección automatizada de noticias falsas en la era de las bibliotecas digitales"*. Y a Oriola (2021) con su publicación "*Exploración de N-gram, incrustación de palabras y modelos de temas para la detección de noticias falsas basadas en contenido en la evaluación de FakeNewsNet".*

PremaLatha et al. (2020) indicaron que el objetivo principal de la automatización robótica de procesos (RPA) es realizar tareas repetitivas sin interacción humana. Podemos ahorrar tiempo, costo, empoderamiento de los empleados y podemos producir los resultados precisos sin ninguna desviación en términos de calidad, brinda el resultado perfecto. RPA es una metodología de desarrollo que utiliza robots de programación para capturar y descifrar aplicaciones existentes para preparar intercambios, controlar información y hablar con otros sistemas de software (p. 4876).

En este sentido Akshay et al. (2020) indicaron que la automatización del cliente de correo electrónico propuesta, es un robot que tiene una capacidad independiente para comunicaciones por correo electrónico efectivas y una funcionalidad reforzada para clasificar de manera inteligente el correo electrónico entrante, en función de la lista de recomendaciones (Akshay et al., 2020, p. 794). El usuario puede escribir en su respuesta que se puede enviar a cada uno de los clientes (ID de correo electrónico del cliente) personalmente que pertenecen a la misma categoría. El sistema también podría incluir la categorización de correos

en función del tema de discusión para optimizar el enrutamiento y la asignación. La inteligencia artificial podría integrarse para generar una respuesta basada en el tema (Akshay et al., 2020, p. 794).

También Borg et al. (2021) tuvieron como objetivo investigar cómo el uso de un clasificador automático basado en aprendizaje automático puede aumentar el rendimiento de la clasificación cuando se trata de clasificar los correos electrónicos entrantes de atención al cliente de una organización. Los clasificadores estudiados se evalúan en un conjunto de datos etiquetado utilizando reglas basadas en palabras clave manejadas manualmente y que actúa como la línea de base en el estudio. (Borg et al., 2021, p. 1882)

Por otro lado, en relación al phishing Al-Alyan et al. (2020) nos mencionaron que los sitios web de phishing intentan hacerse pasar por otros sitios web o entidades para convencer a los usuarios de que ingresen su información personal. Esto hace que los sitios web de phishing sean peligrosos, especialmente para los usuarios aficionados (Al-Alyan et al., 2020, p. 2753). Una solución de detección de phishing de URL que utiliza un modelo de red neuronal convolucional (CNN) a nivel de carácter para clasificar la URL. La CNN aprende de la cadena de URL como una secuencia de caracteres, para evitar características de URL predeterminadas. Por ejemplo, el número de puntos o la longitud de la URL (Al-Alyan et al., 2020, p. 2753).

De la misma forma Dutta indicó que La detección de URL maliciosas se considera una tarea de clasificación binaria con predicciones de dos clases: (a) maliciosa y (b) benigna. El entrenamiento del método ML consiste en encontrar el mejor mapeo entre el espacio vectorial d-dimensional y la variable de salida. Esta estrategia tiene una gran capacidad de generalización para encontrar URL maliciosas desconocidas en comparación con el enfoque de lista negra. (Dutta 2021 p. 2)

Y en relación a las Noticias falsas Mertoğlu et al. (2020) indicaron que, si bien las noticias en formato digital tienden a ser más rápidas, más fáciles, comparativamente más baratas y ofrecen conveniencia en términos de acceso de las personas a la información, apresura la distribución de noticias falsas. Para solucionarlo se necesitó desarrollar un mecanismo automatizados para la variación de la credibilidad del contenido digital servido en bibliotecas o fuentes acreditadas sin necesidad de una interacción humana en el proceso. En ese

sentido se desarrolló un modelo automatizado que detecte las noticias falsas enfocado solo en contenido digitales (Mertoğlu et al., 2020, p. 1).

También Oriola indicó que FakeNewsNet fue motivado por el compromiso social y la conducta de los usuarios en las redes sociales (consumidores de noticias digitales), así como por cómo capturar información dinámica relacionada con la propagación de noticias falsas, las reacciones de los usuarios a las noticias falsas y los patrones temporales para la detección e intervención temprana de noticias falsas (Oriola, 2021, p. 24).

FakeNewsNet es un repositorio de datos público multidimensional, que contiene dos conjuntos de datos con contenido de noticias, contexto social e información dinámica. Según el desarrollador, los conjuntos de características brindan la oportunidad de realizar un estudio exploratorio de diferentes enfoques para una mejor comprensión de las tácticas de desinformación. (Oriola, 2021, p. 24).

Por ultimo tenemos a Soto et al. (2020) que implementa aun método de clasificación para las páginas web tipo phishing usando la minería de datos, también compara varios modelos propuestos como SVM, XGBoost y AdaBoost. Lo cuales registran un uso de CPU y memoria RAM muy eficientes.

En el siguiente apartado se describe las teorías relacionadas. Resaltando el RPA para la clasificación de correos electrónicos, siendo esta la variable de la investigación. También se precisará sobre las dimensiones e indicadores como la reducción del tiempo de clasificación (Nawaz 2019), eficacia de la clasificación (Akshay 2020), uso de CPU (Prexawan-prasut et al., 2017, p. 785; Soto et al., 2020, p.36) y uso de memoria RAM (Soto et al., 2020, p.36; Verona et al., 2016, p. 281).

Akshay y PremaLatha abarcaron el tema de la automatización de procesos robóticos para la clasificación de correos electrónicos (Akshay et al., 2020, p. 794; PremaLatha et al., 2020, p. 4882) y la importa de su aplicación. También los algoritmos de base ***ELM*** (Luo et al., 2021, p. 6) y ***k-step CD-K*** (Yi et al., 2018, p. 3) ver anexo 06.

En relaciona las dimensiones e indicadores del presente estudio Nawaz indicaron que el software RPA se integra con las funciones y herramientas

existentes para manejar tareas básicas a través de la automatización y reducir el consumo de tiempo (Nawaz 2019 p. 609).

Del mismo modo Akshay y compañía indicaron que se observó que es difícil para un usuario revisar todos los correos nuevos que ingresan a su bandeja de entrada, segregarlos e identificar correos electrónicos no deseados. los correos electrónicos que ingresan a la bandeja de entrada se segregarán según la lista de recomendaciones en categorías especificadas (Akshay et al., 2020, p793).

También Iang et al. (2021) habla sobre la clasificar los correos electrónicos en distintas etiquetas puede tener un gran impacto en la atención al cliente. Al utilizar el aprendizaje automático para etiquetar los correos electrónicos, el sistema puede configurar colas que contengan correos electrónicos de una categoría específica. (Iang et al., 2021, p. 1)

Por último, se describe el marco conceptual como la automatización de procesos robóticos por Nawaz (2019), el phishing mencionado por Yi et al. (2018), las noticias falsas por Gutiérrez-Coba et al. (2020), Phishtank descrito por Dutta (2021) y para finalizar el concepto de FakeNewsNet por Oriola (2021).

Nawaz indicó que la automatización de procesos robóticos es un software integrado de tecnología que utiliza los robots para reemplazar las acciones humanas para realizar tareas (Nawaz 2019, p. 609).

Yi et al. (2018) comentaron que el phishing web es el intento de sustraer información sensible o confidencial, así como los nombres, usuarios, contraseñas y tarjetas de crédito (numero, fecha de vencimiento y código de seguridad) esto por motivos maliciosos, haciéndose pasar por un sitio web confiable en Internet (Yi et al., 2018, p. 2).

Gutiérrez-Coba plantío que la definición de noticias falsas (Fake news) ha sido abarcado por varios autores, no obstante, no existe una descripción universal del término. Una parte de los investigadores hacen una distinción entre los términos de desinformación: son noticias elaboradas con la intención de engañar, y también tenemos a misinformation: son noticias diseñada sin la intención de engañar, no obstante, terminan desinformando (Gutiérrez-Coba et al., 2020, p. 238).

Dutta describió el concepto de Phishtank es un conjunto de datos de referencia de sitios web de phishing familiar. Por lo tanto, Phishtank ofrece un conjunto de datos de sitios web de phishing en tiempo real. Los in-vestigadores para establecer la recopilación de datos para probar y detectar sitios web de Phishing utilizan el sitio web de Phishtank (Dutta 2021 p. 5).

Oriola indicó que FakeNewsNet es un repositorio de dos conjuntos de datos novedosos, PolitiFact y GossipCop, que se emplean para la evaluación de técnicas de detección de noticias falsas. A diferencia de otros conjuntos de datos de noticias falsas de referencia ampliamente estudiados, los conjuntos de datos de FakeNewsNet incorporan contenido de noticias, contexto social e información dinámica, que podrían usarse para estudiar la propagación, detección y mitigación de noticias falsas. (Oriola, 2021, p. 24).

III. METODOLOGÍA

En este capítulo del estudio abordaremos el tipo y diseño que se empleara a la investigación. También se desarrollará los siguientes puntos la Variable, Población, muestra y muestreo, Procedimientos, Método de análisis de datos y Aspectos éticos de investigación según el Código de Ética en Investigación de la Universidad Privada César Vallejo

3.1 Tipo y diseño de investigación

Esta investigación es de tipo aplicada, ya que se realizará una automatización robótica de procesos para clasificación de correos que nos lleva al desarrollo de conocimiento práctico para un mejor entendimiento y por ende a la resolución de problemas. Al respecto, Keys y compañía indicaron que uno de los métodos para ayudar a adaptarse a las tecnologías emergentes es introducir ejercicios prácticos para aprenden cómo usar estas tecnologías para resolver problemas (Keys et al., 2020 p. 25).

En ese sentido la investigación tuvo un enfoque cuantitativo esto se debe a que los datos a recopilar deben ser de naturaleza cuantitativo. De tal manare tenemos como sustento lo indicado por Hernández et al. (2016) donde manifiesta que la investigación cuantitativa recopila datos e información de naturaleza cuantitativa y se apoya con la estadística; También menciona que los estudios cuantitativos seguían de una secuencia preestablecida y estructurado (el proceso), se debe tener en cuenta que las decisiones fundamentales sobre el método que se elegirá antes de recopilar los datos (p. 6).

El diseño del estudio fue pre-experimental, porque se probará los algoritmos bases y el propuesto teniendo encueta y se observará el desempeño. Cedeño indicaron que se emplearon como métodos el pre-experimento para la observación al desempeño de los estudiantes (Cedeño et al., 2018 p. 71).

3.2 Variables y operacionalización

Impacto de un Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news. En relación de a la variable Akshay et al. (2020) indicaron una funcionalidad reforzada para clasificar de manera inteligente el correo

electrónico (p. 794). También PremaLatha et al. (2020) comenta que permite a los estudiantes registrar los datos y analizar esos datos y clasificarlos en datos (p. 4882). La matriz de operacionalización de la variable tiene el cambio de las variables teóricas hacia dimensiones e indicadores. La información está en el "*Matriz de Operacionalización*" (ver anexo 02).

3.3 Población, muestra y muestreo

A continuación, se comentará las definiciones relacionada a la población, muestra y muestreo:

A. Población: Estaría compuesta de una gran cantidad de correos seleccionados de la empresa. Chaipornkaew et al. (2017) mencionaron para la investigación inicial, se seleccionaron 12.465 de los correos electrónicos. (Chaipornkaew et al., 2017 p. 2)
B. Muestra: la cantidad o el tamaño de la muestra, y fue determinada por Prexawanprasut et al. (2017) mencionaron que los investigadores seleccionaron 8.000 correos electrónicos (p. 784) desde julio 2015 hasta junio 2016.
C. Muestreo: por conveniencia para que sea más accesible.

3.4 Técnicas e instrumentos de recolección de datos

En esta investigación de RPA para la clasificación de correos se utilizó la ficha de recolección de datos y también la técnica de observación de datos como los instrumentos de medición. En ese sentido Ñaupas et al. (2018) comento que las técnicas de investigación son una compilación de reglas y métodos para regularizar un determinado proceso y también alcanzar un objetivo establecido previamente (p. 13).

Sera necesario tener un instrumento como el monitor de recursos que registre los datos de las herramientas empleadas como uipath, maven, cucumber, Excel, Outlook, Chrome y dependencias de uipath. Las fichas que fueron utilizadas son: el FORM01 donde se registrara la Reducción tiempo de clasificación (ver anexo 03), en el FORM02 se registrara la Eficacia de la clasificación (ver anexo 03), en el FORM03 se indicara el Uso de CPU (ver anexo

03) y en el FORM04 se mostrara el Uso de memoria RAM (ver anexo 03). En todos los formatos se realizar una comparación con los otros modelos desarrollados.

3.5 Procedimientos

Además de implementar el RPA para reportar correos electrónicos sospechosos, es necesario programar un sistema (conjunto de procesos RPA) para analizar estos reportes. Todo el análisis se realiza mediante RPA, junto con algunas operaciones realizadas a través de Base de datos. La siguiente es una representación de alto nivel del proceso de análisis general:

• Inicialización y pre-ejecuciones el robot de clasificación: verificación de la conexión y pre-requisitos de forma interna según la arquitectura empleada. Figura 1 (ver anexo 4).

• Acceder al buzón creado para informar incidentes de phishing/ Fake news y extraer los correos electrónicos potencialmente maliciosos. Figura 2 (ver anexo 5).

• Se realizó una evaluación mediante un algoritmo de para tomar una decisión sobre la clasificación del correo como “Normal”, “Phishing” o “Fake news” Figura 3 (ver anexo 5).

• Se realizó una búsqueda utilizando web scraping para mantener la base de datos actualizada. Figura 4 (ver anexo 5).

• Se guardó la búsqueda en Excel para mantener la base de datos actualizada y se ejecuta una macro predefinida. Figura 5 (ver anexo 5).

• Se realizó una evaluación los correos que contenga URL de forma general, los correo que no cumplan no serán analizados. Figura 6 (ver anexo 5).

• Se realizó una evaluación mediante la búsqueda de URL malicioso de forma específica, dicha información se guardará en una base de datos temporal. Figura 7 (ver anexo 5).

• Se realizó la búsqueda de correos almacenados en la base de datos temporal. Figura 8 (ver anexo 5).

• Se procedió con mover los correos con URL malicioso. Figura 9 (ver anexo 5).

• Cuando haya confirmación de la clasificación del correo electrónico enviara a una bandeja distinta a la principal con el nombre “Phishing_FakeNews”.

En esta fase, el programa primero reordenaría los correos electrónicos en cada categoría. Los correos electrónicos y sus características se recopilaron y almacenaron en la base de datos. Las características del correo electrónico incluían datos del cliente como nombre, dirección, link adjuntas y otros datos. (Prexawanprasut et al., 2017 p. 786).

3.6 Método de análisis de datos

La información recolectada del efecto del uso de la aplicación RPA para la clasificación de correos electrónicos fue ingresada en las tablas mostradas (anexo 3) y se empleó las fórmulas de promedios para poder medir el tiempo de clasificación por cada correo. También se realizar la matriz de confusión con sus componentes (faso negativo, verdadero positivo, verdadero negativos y faso positivo) para calcula la sensibilidad especificidad precisión y exactitud de tal modo que se obtiene la eficiencia en clasificación. Para Uso de CPU y Uso de memoria RAM se capturo la información durante la ejecución del RPA y se empleó las fórmulas de promedios

3.7 Aspectos éticos

En el siguiente apartado se mencionará los aspectos éticos tomados en cuenta para la investigación. Por el lado de la información obtenida en los artículos científicos, patentes, libros y documentos válidos que dan soporte como fuentes de investigación, se brinda el crédito a los autores con su debida referencia bibliográfica, citas y paráfrasis con las normas internacionales APA séptima edición.

Este trabajo de estudio muestra originalidad y un correcto citado de las referencias empleada en la investigación. En cumplimiento de lo declarado en el 9° artículo del código de ética en investigación (2020) de la Universidad César Vallejo, el cual nos informa de la política anti-pagio en los estudios. En el manual de ética (2020) también indico lo siguiente La Universidad César Vallejo alienta

la originalidad e innovación en las investigaciones científicas. Por otro lado, el plagio es el delito que consiste en hacer pasar como de autoría propia un trabajo, proyecto, obra y/o idea no concebida por uno mismo, ya fuese de manera parcial o total (p9).

La Universidad César Vallejo alienta el respeto a los derechos de autor y sanciona severamente a aquellos autores o coautores que realicen plagio (parcia/total) o algún conducto que este fuera de los lineamientos éticos de la investigación.

IV. RESULTADOS

En el presente capítulo se validará el efecto del uso de una automatización robótica de procesos para la clasificación de correos electrónicos según los indicadores como: el tiempo de clasificación, la eficacia de la clasificación, uso de CPU y uso de memoria RAM. De la misma forma se demostró el incremento la eficacia de la clasificación y reducción del tiempo de clasificación de los correos; no obstante, no se produjo una reducción de uso de CPU y uso de memoria RAM en relación a otros modelos

IV.1 Datos descriptivos

En este análisis, se detalla los indicadores utilizado en la investigación. En relación al indicador de Reducción del tiempo, se tomó los tiempos de ejecución y la cantidad de correos analizados por el RPA. En relación al indicador de Incremento la eficacia, se realizó matriz de confusión donde se indica lo real vs lo predictivo por el RPA. Y para el Indicador de Reducción de uso de CPU y reducción de uso de memoria RAM, se monitorio el porcentaje y la cantidad en megabyte (MB) que utiliza el RPA al momento de clasificar los correos.

IV.1.1 Indicador de Reducción del tiempo de clasificación de correos phishing y fake news

En la siguiente tabla nos muestra el detalle de la “cantidad de correos electrónicos” y el “tiempo de ejecución” que el RPA empleo en analizar cada grupo de correo, para luego realizar un promedio del tiempo por cada correo

Tabla 1: Cantidad de correos y tiempo de ejecución

Cantidad correo	Tiempo de ejecución Seg.	Segundos/Correos
500	228	0.46
1000	346	0.35
1100	368	0.33
1500	468	0.31
125	10	0.08
300	161	0.54
1600	470	0.29
1414	427	0.30
1700	751	0.44
2000	532	0.27
	$T_{prom.}$	**0.34**

IV.1.2 Indicador de Incremento la eficacia de la clasificación de correos phishing y fake news

En la siguiente tabla nos muestra el detalle de lo “estimado por el modelo” versus lo “real” luego de la clasificación de correos para realizar la matriz de confusión, esto nos permitió determinar los valores verdadero positivo, verdadero negativo, falso negativo y falso positivo

Tabla 2: Eficacia de la clasificación

Matriz de confusión		**Estimado por el Modelo**		
		Positivo (P)	Negativo (N)	**Total de correo**
Real	Verdadero (V)	4928	0	4928
	Falso (F)	0	3072	3072
	Total de correo	4928	3072	8000

IV.1.3 Indicador de Reducción de Uso de CPU en la clasificación de correos phishing y fake news

En la siguiente tabla nos muestra el detalle del uso de CPU durante las ejecuciones de clasificación. También se muestra un promedio de dicho uso

Tabla 3: Uso de CPU durante las ejecuciones

Nº	Uso Promedio en %
1	0.04
2	0.111666667
3	0.644444444
4	0.345
5	0.544615385
6	0.287692308
7	0.51
8	0.2725
9	0.03
10	0.171111111
$U_{prom.}$	**0.2957**

IV.1.4 Indicador de Reducción de Uso de memoria RAM en la clasificación de correos phishing y fake news

En la siguiente tabla nos muestra el detalle del uso de memoria RAM durante las ejecuciones de clasificación. También se muestra un promedio de dicho uso

Tabla 4: Uso de memoria RAM durante las ejecuciones

Nº	Uso Promedio en MB
1	309.50
2	302.82
3	259.79
4	324.07
5	276.17
6	261.47
7	259.00
8	306.67
9	297.05
10	267.06
$T_{prom.}$	**286.36**

IV.2 Prueba de hipótesis

Para la validación de la hipótesis HE2 (Indicador de incremento la eficacia de la clasificación de correos phishing y fake news) se realizó la matriz de confusión y se comparó con otros modelos. Mientas que para validar la hipótesis HE1 (Indicador de reducción del tiempo de clasificación de correos phishing y fake news), HE3 (Indicador de reducción de uso de CPU en la clasificación de correos phishing y fake news) y HE4 (Indicador de reducción de Uso de memoria RAM en la clasificación de correos phishing y fake news) se realizó una comparación de los promedios versus otros modelos.

IV.2.1 Hipótesis específica 1

$HU1_0$ El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K no redujo el tiempo de clasificación

$HU1_1$ El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el tiempo de clasificación

Tabla 5: Promedio de cantidad de correos y tiempo de ejecución

Modelos	Tiempo
ELM - K-step CD-K	**0.34**
ELM	0.77
LSTM	2.00
Ramdom Forest	0.98

IV.2.2 Hipótesis específica 2

$HE2_0$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K no incremento la eficacia de la clasificación.

$HE2_1$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K incremento la eficacia de la clasificación.

Tabla 6: Promedio de eficacia de la clasificación

Indicador:		Incremento la eficacia de la clasificación de correos phishing y fake news			
Nº	Modelos	Sensibilidad	Especificidad	Precisión	Exactitud
1	**Maximum Entropy**	99.60%	83.50%	85.79%	91.55%
2	**AdaB-ForestPA-PWDM**	98.08%	96.55%	97.28%	97.40%
3	**LSTM**	90.25%	-	91.00%	99.00%
4	**Ramdom Forest**	99.84%	87.50%	98.04%	98.14%
5	**ELM - K-step CD-K**	100.00%	100.00%	100.00%	100.00%

IV.2.3 Hipótesis específica 3

$HE3_0$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K no redujo el uso de CPU

$HE3_1$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de CPU

Tabla 6: Promedio de uso del CPU durante las ejecuciones

Modelos	Uso en %
ELM - K-step CD-K	29.57%
SVM	17.64%
XGBoost	21.71%
AdaBoost	43.17%

IV.2.4 Hipótesis específica 4

$HE4_0$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K no redujo el uso de RAM

$HE4_1$: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de RAM

Tabla 7: Promedio uso de memoria RAM durante las ejecuciones

Modelos	Uso en MB
ELM - K-step CD-K	**286.36**
SVM	15.50
XGBoost	4.00
AdaBoost	409.00

IV.2.5 Hipótesis general

En este apartado se detalla de forma abreviada los resultados para la validación de la hipótesis general (HG).

HG_0: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K no redujo el tiempo de clasificación, no incremento la eficacia de la clasificación*, no redujo el uso de CPU y no redujo el uso de RAM*

HG_1: El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el tiempo de clasificación, incremento la eficacia de la clasificación, redujo el uso de CPU y redujo el uso de RAM

Tabla 8: Indicadores de las hipótesis de la investigación

	ELM - K-step CD-K	Modelo de comparación	Valor del indicador del modelo de comparación
Tiempo de clasificación	**0.34 s**	ELM	0.77 s
Eficacia de la clasificación	**100.00%**	LSTM	99.00%
Uso de CPU	29.57%	SVM	**17.64%**
Uso de memoria RAM	286.36 MB	XGBoost	**4.00 MB**

En la tabla 8, se puede observar los promedios y resultados de cada indicador de la investigación. De tal forma se afirma que la automatización robótica de procesos para la clasificación de correos electrónicos redujo el tiempo de clasificación, incremento la eficacia de la clasificación, no redujo el uso de CPU y no redujo el uso de memoria RAM. En ese sentido, se acepta la hipótesis alternativa y se rechazó la hipótesis nula. En la siguiente tabla se muestra el resumen de resultados de las hipótesis.

Tabla 9: Resumen general de los resultados de comprobación de las hipótesis.

	Detalle de la Hipótesis	Resultado
HE1	El uso de la aplicación RPA para la clasificación de correos redujo el tiempo de clasificación	Aceptación
HE2	El uso de la aplicación RPA para la clasificación de correos incremento la eficacia de la clasificación	Aceptación
HE3	El uso de la aplicación RPA para la clasificación de correos redujo el uso de CPU	Rechazada
HE4	El uso de la aplicación RPA para la clasificación de correos redujo el uso de la memoria RAM	Rechazada
HG	El uso de la aplicación RPA para la clasificación de correos redujo el tiempo de clasificación, incremento la eficacia de la clasificación, redujo el uso de CPU y redujo el uso de RAM	Rechazada

V. DISCUSIÓN

En el presente capítulo se mostrará un resumen de los resultados más importantes de la investigación, los cuales serán comparados con las investigaciones pasadas y con los artículos científicos existentes en relación al tema. En el siguiente párrafo se detalló los resultados que se produjeron al validar la hipótesis general, después se detalló cada indicador de forma individual con su resultado obtenido versus los valores obtenido en otros artículos científicos o investigaciones de tal modo se genera la discusión con naturaleza científica.

El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K obtuvo un tiempo de clasificación de 0.34 en segundos, lo que fue menor al resultado del modelo ELM que obtuvo 0.77 segundos (Montoya et al., 2020). Del mismo modo, se obtuvo un porcentaje de eficacia de 100%, lo cual es mayor al 99% planteado por el modelo LSTM (Borg et al., 2021). Asimismo, el porcentaje de uso del CPU usado por la aplicación RPA fue 29.5703%, mientras que en el modelo de SVM fue 17.64% (Soto et al., 2020). De igual manera, el uso de memoria RAM usado por la aplicación RPA fue 286.36 MB, mientras que en el modelo de XGBoost fue de 4.00 MB (Soto et al., 2020).

En esta investigación se obtuvo un tiempo de clasificación de correos como phishing o fake news de 0.34 segundos, con una aplicación RPA basada en el algoritmo fusionado ELM - K-step CD-K que analiza el contenido del correo, el que fue menor a los tiempos de clasificación de estudios previos por las siguientes razones: (a) se asegura que el cuerpo no tenga URL, (b) solo hay análisis detallado si el cuerpo tiene URL y (c) el análisis del URL es directo con la base de datos.

Mientras Montoya et al. (2020) se obtiene 0.77 s como tiempo con el modelo ELM porque el flujo contiene muchos condicionales. Borg et al. (2021) se obtiene un tiempo de clasificación de 2.00 segundos con el modelo de LSTM con Deep learning classifiers requiriendo más tiempo de análisis. Céspedes (2021) con el modelo de Ramdom Forest con tiempo de clasificación de 0.98 segundos en encontrar URLs maliciosas con machine learning.

En relación al incremento de eficacia se tiene en cuenta los parámetros de sensibilidad, especificidad, precisión y exactitud. Con el modelo Maximum Entropy se tiene el 99.60% de sensibilidad (Asani et al., 2021) utilizando un analizador por palabras enfocado al correo. El modelo AdaB-ForestPA-PWDM se tiene el 96.55% de especificidad (Alsariera et al., 2020) utilizando machine learning enfocado a los sitios web phishing. El modelo Ramdom Forest se tiene el 98.04% de precisión (Céspedes, 2021) utilizando machine learning enfocado al URLs maliciosas. El modelo LSTM se tiene el 99.00% de exactitud (Borg et al., 2021) con Deep learning classifiers.

Mientras el modelo propuesto ELM - K-step CD-K obtuvo un 100% en los parámetros de sensibilidad, especificidad, precisión y exactitud, resaltando una mejor eficacia ya que el análisis del URL es directo con la base de datos y se hace un mejor descarte de los correos que no contengan URL disminuyendo a cero la cantidad de “falso negativo” y “falso positivo”.

En relación al uso del CPU se obtuvo un resultado del 29.57% el cual se debe a la implementación de una automatización robótica de procesos para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K esto se debe al uso de otros programas como Excel, Outlook y Chrome.

Mientras Soto et al. (2020) obtubieron como porcentaje de uso del CPU un 17.64% con el modelo de SVM esto evidencia un mejor desempeño en CPU ya que utiliza menor recurso. El modelo de XGBoost obtuvo un 21.71% de uso del CPU. El modelo que obtuvo más uso del CPU fue AdaBoost con un 43.17%. Estos modelos no utilizan los programas de office como soporte (Soto et al., 2020).

En relación al uso de la memoria RAM se obtuvo un resultado del 286.36 MB el cual se debe a la implementación de una automatización robótica de procesos para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K esto se debe al uso de otros programas como Excel, Outlook y Chrome. El modelo que obtuvo más uso de la memoria RAM fue AdaBoost con un 409.00 MB ya que tiene un análisis más robusto (Soto et al., 2020) y requiere más recursos.

Mientras en la investigación de Soto et al. (2020) se obtiene como uso de la memoria RAM un 15.50 MB con el modelo de SVM. Por otro lado, el modelo de XGBoost obtuvo 4.00 MB de uso de la memoria RAM esto evidencia un mejor desempeño en memoria RAM ya que utiliza menor recurso. Estos modelos no utilizan los programas de office como soporte. Estos modelos no utilizan otros programas de soporte y su automatización está bien definida por su árbol de decisiones (Soto et al., 2020).

Conforme a los indicado con anterioridad, se puede llegar a la conclusión que la automatización robótica de procesos para la clasificación de correos electrónicos como phishing y fake news, tuvo como resultados un buen desempeño. De tal modo se evidencia los resultados obtenidos, cumplen con el mejor tiempo de clasificación y una excelente eficiencia del efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news detallado por Akshay et al. (2020) y PremaLatha et al. (2020). Por ultimo en relación al uso de CPU y de memoria RAM, el RPA propuesto no cumplió con supera a los modelos propuestos en investigaciones pasadas.

VI. CONCLUSIONES

En este capítulo se detallará las conclusiones de la investigación:

1. Se obtuvo un tiempo de clasificación de correos como phishing o fake news de 0.34 segundos. Se concluyó que el RPA propuesto tiene mejor desempeño según el objetivo específico de tiempo de clasificación, ya que se asegura que el cuerpo no tenga URL, también solo hay análisis detallado si el cuerpo tiene URL y por último el análisis del URL es directo con la base de datos.

2. En tiempo de clasificación Montoya et al. (2020) obtuvieron 0.77 s con el modelo ELM, Borg et al. (2021) obtuvieron 2.00 segundos con el modelo de LSTM y Céspedes (2021) con el modelo Ramdom Forest se obtuvo 0.98 segundos. Se concluyó que estos modelos tienen un análisis más complejo por ende el incremento del tiempo.

3. En la eficacia se tiene en cuenta los parámetros de sensibilidad, especificidad, precisión y exactitud. Con el modelo Maximum Entropy se tiene el 99.60% de sensibilidad (Asani et al., 2021). El modelo AdaB-ForestPA-PWDM se tiene el 96.55% de especificidad (Alsariera et al., 2020). El modelo Ramdom Forest se tiene el 98.04% de precisión (Céspedes, 2021). El modelo LSTM se tiene el 99.00% de exactitud (Borg et al., 2021). Se concluyó que los modelos tienen un margen de error en precisar “falso negativo” y “falso positivo”.

4. El modelo ELM - K-step CD-K obtuvo un 100% en los parámetros de sensibilidad, especificidad, precisión y exactitud, se concluyó que el RPA propuesto tiene mejor desempeño según el objetivo específico de eficacia, ya que el análisis del URL es directo con la base de datos y se hace un mejor descarte de los correos que no contengan URL disminuyendo a cero la cantidad de “falso negativo” y “falso positivo”

5. En relación al uso del CPU se obtuvo un resultado del 29.57% el cual se debe a la implementación de una automatización robótica de procesos para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K se

concluyó que los soportes de distintos programas hacen un incremento en el uso de CPU.

6. El modelo de XGBoost obtuvo un 21.71% de uso del CPU. El modelo de AdaBoost obtuvo un 43.17% de uso del CPU. El modelo de SVM obtuvo un 17.64% de uso del CPU. si bien estos modelos no utilizan los programas de office como soporte, se concluyó que el modelo SVM cumple con el objetivo específico de uso de CPU ya que utiliza menos recursos. (Soto et al., 2020).

7. En relación al uso de la memoria RAM se obtuvo un resultado del 286.36 MB el cual se debe a la implementación de una automatización robótica de procesos para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K que los soportes de distintos programas. El modelo AdaBoost su uso de la memoria RAM fue de 409.00 MB ya que tiene un análisis más robusto (Soto et al., 2020) se concluyó que estos modelos hacen un uso elevado de RAM.

8. Soto et al. (2020) el modelo de SVM obtuvo 15.50 MB de uso de RAM. El modelo de XGBoost obtuvo 4.00 MB de uso de RAM. Estos modelos no utilizan los programas de office como soporte y su automatización está bien definida por su árbol de decisiones. Se concluyó que XGBoost cumple mejor con el objetivo específico de uso de RAM (Soto et al., 2020).

9. En resumen, la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K se desempeñó de forma positiva, Se infiere que el mejor tiempo de clasificación es de 0.34 ya que se asegura no se pierde tiempo analizado a correos que no contenga URL. También, se concluyó que el RPA propuesto tiene mejor eficiencia obteniendo un 100% ya que el análisis es directo con la base de datos.

10. Asimismo, el porcentaje de uso del CPU usado por la aplicación RPA fue de 29.5703%, mientras que en el modelo de SVM fue de 17.64% y de igual manera, el uso de memoria RAM usado por la aplicación RPA fue de 286.36 MB, mientras que en el modelo de XGBoost fue de 4.00 MB en ambos indicadores se concluyó que los modelos SVM y XGBoost no requieren otros programas de soporte de tal modo consumen menos recursos.

VII. RECOMENDACIONES

En este capítulo se indicará las recomendaciones para futuras investigaciones:

1. Emplear otras herramientas de automatización como Serenity o Selenium para una futura comparación con Uipath. Así tener otros parámetros de variación como correos corporativos y gratuitos, máxima cantidad de correos a mover, mejor compatibilidad en plataformas o sistemas operativos y por último los costos de inversión.

2. Considerar utilizar nuevos indicadores para incrementar el conocimiento y diversificar la llegada a los usuarios finales, los cuales puede ser: facilidad de instalación, tamaño de archivos generado y/o adaptación al usuario final.

3. Crear nuevas carpetas de categorización usando la automatización robótica de procesos para clasificar de manera inteligente (Akshay et al. 2020, p. 794). Con ese mismo sentido se puede agregar las categorías como área-laboral puesto donde labora el usuario, también área externa-laboral, Clientes, proveedores y otras clasificaciones según el puesto que desempeña el usuario. Mejorando la experiencia al usuario

4. Aplicar las categorización o clasificación de los correos en un entorno Web, para que puede ser utilizado por Outlook, Yahoo! Mail y Gmail. En ese sentido se puede proponer “una interfaz basada en web” (Borg et al., 2021), lo cual puede llevar el RPA propuesto a masificarse en varias organizaciones y en distintos países.

5. Para el apartado de Noticias falsas se puede crear un analizador de contenido regional (Latinoamérica) o hasta Local (Perú) usando Web scraping con los principales sitios web de noticias que tenga respaldo y credibilidad, para una mejor discriminación de las noticias que no tienen fundamente verídico o con información ambigua.

6. Siguiendo con el apartado de Noticias falsa se puede proponer una RPA para la Redes Sociales (Posadas-Durán et al., 2019). Solo en Facebook que es la red social predominante (Salinas 2022) con 2.320 millones de

usuarios esto nos garantiza una gran cantidad de información para su análisis. Del mismo modo también se puede analizar distintos tipos de datos, como URL compartidas, comentarios y publicaciones. Con ello podemos ayudar a disminuir la cantidad de notificas falsas que llegan a las redes sociales.

7. Para finalizar un análisis directo y amplio de los URL enviado con Phishing, se puede utilizar la Red neuronal convolucional (Mourtaji et al., 2021). Esto nos puede ayudar a tener un modelo de RPA combinado con inteligencia artificial y también podemos tener una mejor base de datos de las direcciones maliciosas.

REFERENCIAS

Akshay, P. N., Kalagi, N., Shetty, D., & Ramalingam, H. M. (2020). EMAIL CLIENT AUTOMATION WITH RPA.
Al-Alyan, A., & Al-Ahmadi, S. (2020). Robust URL Phishing Detection Based on Deep Learning. KSII Transactions on Internet & Information Systems, 14(7), 2752–2768. https://doi.org/10.3837/tiis.2020.07.001
Alsariera, Y. A., Elijah, A. V., & Balogun, A. O. (2020). Phishing Website Detection: Forest by Penalizing Attributes Algorithm and Its Enhanced Variations. Arabian Journal for Science & Engineering (Springer Science & Business Media B.V.), 45(12), 10459–10470. https://doi.org/10.1007/s13369-020-04802-1
Asani, E. O., Omotosho, A., Danquah, P. A., Ayoola, J. A., Ayegba, P. O., & Longe, O. B. (2021). A maximum entropy classification scheme for phishing detection using parsimonious features. Telkomnika, 19(5), 1707–1714. https://doi.org/10.12928/TELKOMNIKA.v19i5.15981
Borg, A., Boldt, M., Rosander, O., & Ahlstrand, J. (2021). E-mail classification with machine learning and word embeddings for improved customer support. Neural Computing & Applications, 33(6), 1881–1902.
Céspedes Maestre, M. (2021). Detección de URLs maliciosas por medio de técnicas de aprendizaje automático. Universidad Nacional de Colombia.
Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. PLoS ONE, 16(10), 1–17. https://doi.org/10.1371/journal.pone.0258361
FLYGARE, R. y HOLMQVIST, A. Performance characteristics between monolithic and microservice-based systems. Tesis de Pregrado. Faculty of Computing Blekinge Institute of Technology. Suecia, 2017.
Gutiérrez-Coba, L. M., Coba-Gutiérrez, P., & Gómez-Díaz, J. A. (2020). Noticias falsas y desinformación sobre el Covid-19: análisis comparativo de seis países iberoamericanos. Revista Latina de Comunicación Social, 78, 237–264. https://doi.org/10.4185/RLCS-2020-1476
HERNÁNDEZ, R., FERNÁNDEZ, C. y BAPTISTA, P. Metodología de la investigación. 6. México: McGraw-Hill, 2016. ISBN: 978-1-4562-2396-0.
IANG CHEN, JIAYING PENG, YANG LIU, JINTANG LI, FENFANG XIE, & ZIBIN ZHENG. (2021). Phishing Scams Detection in Ethereum Transaction Network. ACM Transactions on Internet Technology, 21(1), 1–16. https://doi.org/10.1145/3398071
Keys, B., & Zhang, Y. (James). (2020). Introducing RPA in an Undergraduate AIS Course: Three RPA Exercises on Process Automations in Accounting. Journal of Emerging Technologies in Accounting, 17(2), 25–30. https://doi.org/10.2308/JETA-2020-033
Luo, H., & Luo, H. (2021). RPA and Artificial Intelligence in Budget Management Based on Multiperspective Recognition Based on Network Communication Integration. Wireless Communications & Mobile Computing, 1–13. https://doi.org/10.1155/2021/9723379
Mertoğlu, U., & Genç, B. (2020). Automated Fake News Detection in the Age of Digital Libraries. Information Technology & Libraries, 39(4), 1–19. https://doi.org/10.6017/ital.v39i4.12483
Montoya Suárez, L. M., Restrepo Sierra, J., & Gómez Marín, E. (2020). Creación de pacientes internacionales automatizado con AutoIt: un caso de aplicación. Lámpsakos, 24, 74–81. https://doi.org/10.21501/21454086.2809

Mourtaji, Y., Bouhorma, M., Alghazzawi, D., Aldabbagh, G., & Alghamdi, A. (2021). Hybrid Rule-Based Solution for Phishing URL Detection Using Convolutional Neural Network. Wireless Communications & Mobile Computing, 1–24. https://doi.org/10.1155/2021/8241104

Nawaz, D. N. (2019). Robotic process automation for recruitment process. International Journal of Advanced Research in Engineering and Technology, 10(2).

Nayak, P. P., Shrivastava, N., Udham, P. K., Archana, Jadav, R., Chechi, T. S., & Mullasseri, S. (2019). Fake News Detection: Automating the process. Current Science (00113891), 117(11), 1773. https://doi.org/10.1016/j.cogsys.2019.07.004

ÑAUPAS, H., PALACIOS, J., ROMERO, H. y VALDIVIA, M. Metodología y diseños en investigación científica. Cuantitativa–Cualitativa y Redacción de la Tesis. 5. Bogotá: Ediciones de la U, 2018. ISBN: 978-958-762-876-0

ORIOLA, Oluwafemi. Exploring N-gram, word embedding and topic models for content-based fake news detection in FakeNewsNet evaluation. Int J Comput Appl, 2021, vol. 975, p. 8887.

Posadas-Durán, J.-P., Gómez-Adorno, H., Sidorov, G., Escobar, J. J. M., Pinto, D., & Singh, V. (2019). Detection of fake news in a new corpus for the Spanish language. Journal of Intelligent & Fuzzy Systems, 36(5), 4869–4876. https://doi.org/10.3233/JIFS-179034

PremaLatha, V., Krishna, B. M., Kumar, B. N., & Rajesh, T. V. (2020).Email Automation Using Robotic Process Automation (RPA).

Prexawanprasut, T., & Chaipornkaew, P. (2017). Email Classification Model for Workflow Management Systems. Walailak Journal of Science & Technology, 14(10), 783–790.

Piyanuch Chaipornkaew, Takorn Prexawanprasut, Chia-Lin Chang, & McAleer, M. (2017). A Generalized Email Classification System for Workflow Analysis. Journal of Management Information & Decision Sciences, 20, 1–12.

Salinas Arguello, M. X. (2022). Las redes sociales (Bachelor's thesis, Babahoyo: UTB-FAFI. 2022)

Soto, J., Maguiña, J. (2020). Implementación de un método de clasificación de minería de datos para detectar páginas web de tipo phishing [Tesis, Universidad Señor de Sipán]. https://hdl.handle.net/20.500.12802/8895

VERONA, S., PÉREZ, Y., TORRES, L., DELGADO, M. y YÁÑEZ, C. Pruebas de rendimiento a componentes de software utilizando programación orientada a aspectos. Ingeniería Industrial, 2016, 37 (3), pp. 278-285

Yi, P., Guan, Y., Zou, F., Yao, Y., Wang, W., & Zhu, T. (2018). Web Phishing Detection Using a Deep Learning Framework. Wireless Communications & Mobile Computing, 1–9. https://doi.org/10.1155/2018/4678746

Anexo 1: Matriz de operacionalización de variables

Matriz de operacionalización de variables

Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Escala de medición
Efecto de RPA para la clasificación de correos electrónicos	Una funcionalidad reforzada para clasificar de manera inteligente el correo electrónico (Akshay et al., 2020, p. 794) Permite a los estudiantes registrar los datos y analizar esos datos y clasificarlos en datos (PremaLatha et al., 2020, p. 4882)	Resultados del RPA para la clasificación de correos electrónicos	tiempo de clasificación (Akshay et al., 2020, p. 788; Prema-Latha et al., 2020 p. 4876)	Reducción del tiempo de clasificación (Akshay et al., 2020, p. 788; Prema-Latha et al., 2020 p. 4876)	De razón
			eficacia de la clasificación (Akshay et al., 2020, p. 793; Prexawanprasut et al., 2017 p. 784)	Incrementa la eficacia de la clasificación (Akshay et al., 2020, p. 793; Prexawanprasut et al., 2017 p. 784)	De razón
			Uso de CPU (Prexawanprasut et al., 2017, p. 785; Soto et al., 2020, p.36)	Reducción de Uso de CPU (Prexawanprasut et al., 2017, p. 785; Soto et al., 2020, p.36)	De razón
			Uso de memoria RAM (Soto et al., 2020, p.36 ; Verona et al., 2016, p. 281)	Reducción de Uso de memoria RAM (Soto et al., 2020, p.36 ; Verona et al., 2016, p. 281)	De razón

Anexo 2: Matriz de consistencia

Título de la tesis: “Aplicación RPA para la clasificación de correos electrónicos como **phishing** y **fake news**”

PROBLEMAS	OBJETIVOS	HIPOTESIS	VARIABLE	DIMENSIO-NES	INDICADORES
General	**General**	**General**			
¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K?	Determinar el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K	Se validara el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K (Akshay et al., 2020, p. 794; PremaLatha et al., 2020, p. 4882)			
Especifico	**Especifico**	**Especifico**			**Indicadores**
¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el tiempo de clasificación?	Determinar el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el tiempo de clasificación	El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el tiempo de clasificación.	RPA para la clasificación de correos electrónicos (Akshay et al., 2020, p. 794; PremaLatha et al., 2020, p. 4882)	Tiempo de clasificación (Akshay et al., 2020, p. 788; PremaLatha et al., 2020 p. 4876)	Reducción del tiempo de clasificación (Akshay et al., 2020, p. 788; PremaLatha et al., 2020 p. 4876)
¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en la eficacia de la clasificación?	Establecer cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en la eficacia de la clasificación..	El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K incrementa la eficacia de la clasificación.		Eficacia de la clasificación (Akshay et al., 2020, p. 793; Prexawanprasut et al., 2017 p. 784)	Incrementa la eficacia de la clasificación (Akshay et al., 2020, p. 793; Prexawanprasut et al., 2017 p. 784)

¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el Uso de CPU?	Estimar cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el uso de CPU.	El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de CPU.		Uso de CPU (Prexawanprasut et al., 2017, p. 785; Soto et al., 2020, p.36)	Reducción de uso de CPU (Prexawanprasut et al., 2017, p. 785; Soto et al., 2020, p.36)
¿Cuál fue el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phising y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el Uso de memoria RAM?	Medir cuál será el efecto del uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K en el uso de RAM.	El uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de RAM.		Uso de memoria RAM (Soto et al., 2020, p.36 ; Verona et al., 2016, p. 281)	Reducción de uso de memoria RAM (Soto et al., 2020, p.36 ; Verona et al., 2016, p. 281)

Anexo 3: Ficha de recolección de datos

En la siguiente tabla nos nuestra el uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el tiempo de clasificación

Tabla 10: FORM1 Tiempo de clasificación

Cantidad correo	Tiempo de ejecución Seg.	Segundos/Correos
500	228	0.46
1000	346	0.35
1100	368	0.33
1500	468	0.31
125	10	0.08
300	161	0.54
1600	470	0.29
1414	427	0.30
1700	751	0.44
2000	532	0.27
	$T_{prom.}$	0.34

Tabla 11: FORM2 Tiempo de clasificación y comparación de promedios

Ficha de recolección de datos			
Título de la investigación		Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news	
Investigador:		George Eduard Icochea Rivera	
Fecha de recolección de datos:		9/07/2022	
Indicador:		Reducción del tiempo de clasificación de correos phishing y fake news	
Nº	Análisis de Correo/segundo	**Modelos**	**Tiempo**
1	0.46	**ELM - K-step CD-K**	**0.34**
2	0.35	**ELM**	**0.77**
3	0.33	**LSTM**	**2.00**
4	0.31	**Ramdom Forest**	**0.98**
5	0.08		
6	0.54		
7	0.29		
8	0.30		
9	0.44		
10	0.27		
X	**0.34**		

En la siguiente tabla nos nuestra el uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K incremento la eficacia de la clasificación

Tabla 12: FORM3 Eficacia de la clasificación

Matriz de confusión		**Estimado por el Modelo**		
		Positivo (P)	Negativo (N)	**Total de correo**
Real	Verdadero (V)	4928	0	4928
	Falso (F)	0	3072	3072
	Total de correo	4928	3072	8000

Tabla 13: FORM4 Eficacia de la clasificación y comparación de promedios

Ficha de recolección de datos					
Título de la investigación		Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news			
Investigador:		George Eduard Icochea Rivera			
Fecha de recolección de datos:		9/07/2022			
Indicador:		Incremento la eficacia de la clasificación de correos phishing y fake news			
Nº	Modelos	Sensibilidad	Especificidad	Precisión	Exactitud
1	**Maximum Entropy**	99.60%	83.50%	85.79%	91.55%
2	**AdaB-ForestPA-PWDM**	98.08%	96.55%	97.28%	97.40%
3	**LSTM**	90.25%	-	91.00%	99.00%
4	**Ramdom Forest**	99.84%	87.50%	98.04%	98.14%
5	**ELM - K-step CD-K**	100.00%	100.00%	100.00%	100.00%

En la siguiente tabla nos nuestra el uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de CPU.

Tabla 14: FORM5 Uso de CPU y comparación de promedios.

Ficha de recolección de datos	
Título de la investigación	Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news
Investigador:	George Eduard Icochea Rivera
Fecha de recolección de datos:	9/07/2022
Indicador:	Reducción de Uso de CPU en la clasificación de correos phishing y fake news

Nº	Uso Promedio en %	**Modelos**	**Uso en %**
1	0.04	**ELM - K-step CD-K**	**29.57%**
2	0.111666667	**SVM**	**17.64%**
3	0.644444444	**XGBoost**	**21.71%**
4	0.345	**AdaBoost**	**43.17%**
5	0.544615385		
6	0.287692308		
7	0.51		
8	0.2725		
9	0.03		
10	0.171111111		
T prom.	**29.57%**		

En la siguiente tabla nos nuestra el uso de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news basada en el algoritmo fusionado ELM - K-step CD-K redujo el uso de RAM.

Tabla 15: FORM6 Uso de memoria RAM y comparación de promedios

Ficha de recolección de datos	
Título de la investigación	Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news
Investigador:	George Eduard Icochea Rivera
Fecha de recolección de datos:	9/07/2022
Indicador:	Reducción de Uso de memoria RAM en la clasificación de correos phishing y fake news

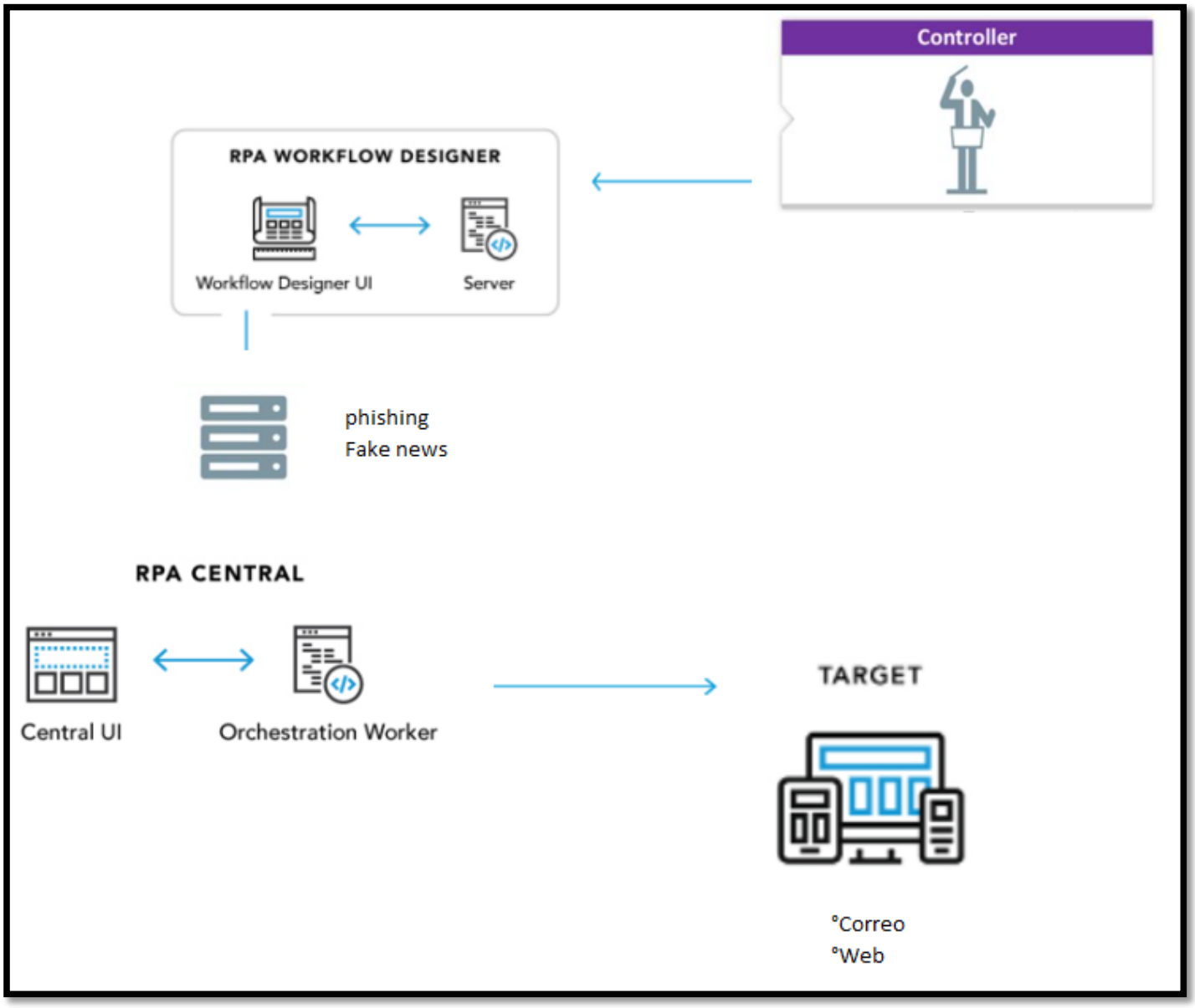
Nº	Uso Promedio en MB	**Modelos**	**Uso en MB**
1	309.50	**ELM - K-step CD-K**	**286.36**
2	302.82	**SVM**	**15.50**
3	259.79	**XGBoost**	**4.00**
4	324.07	**AdaBoost**	**409.00**
5	276.17		
6	261.47		
7	259.00		
8	306.67		
9	297.05		
10	267.06		
T prom.	**286.36**		

Anexo 4: Arquitectura

En la siguiente figura nos nuestra la arquitectura de la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 1 Arquitectura

Fuente: slideplayer.es y sdos.es

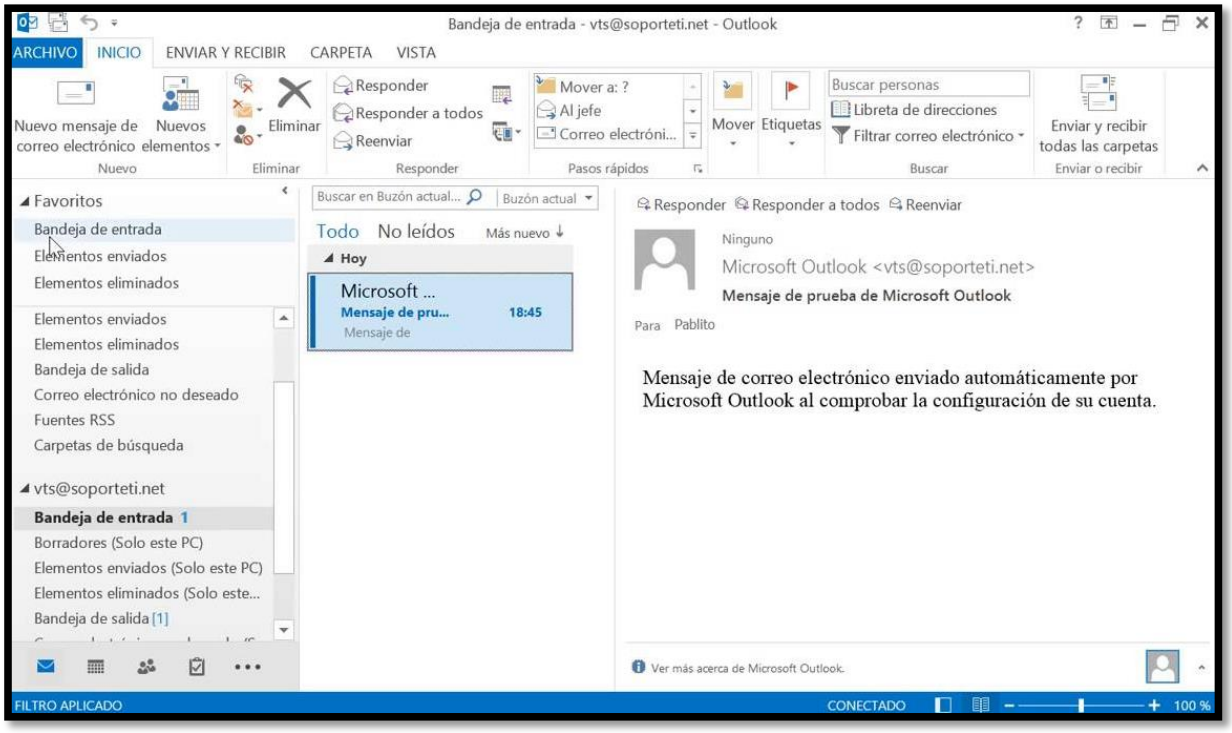


Anexo 1: Prototipos

En la siguiente figura nos nuestra como accede al buzón principal con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 2: Acceder al buzón principal

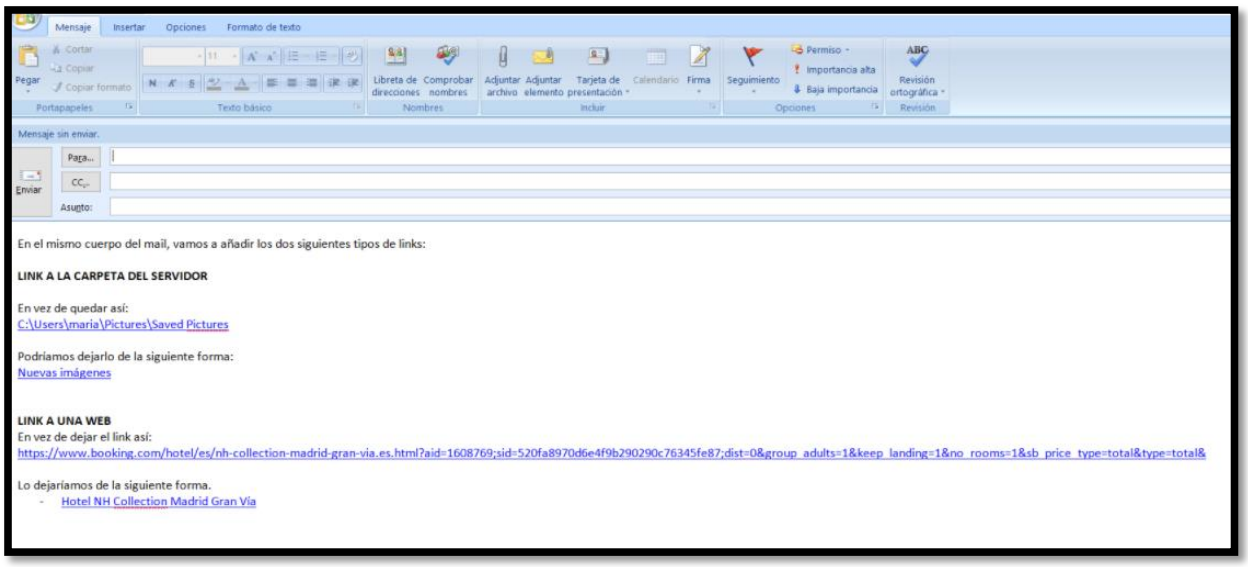
Fuente: i.ytimg.com



En la siguiente figura nos nuestra una evaluación mediante un algoritmo de para tomar una decisión con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

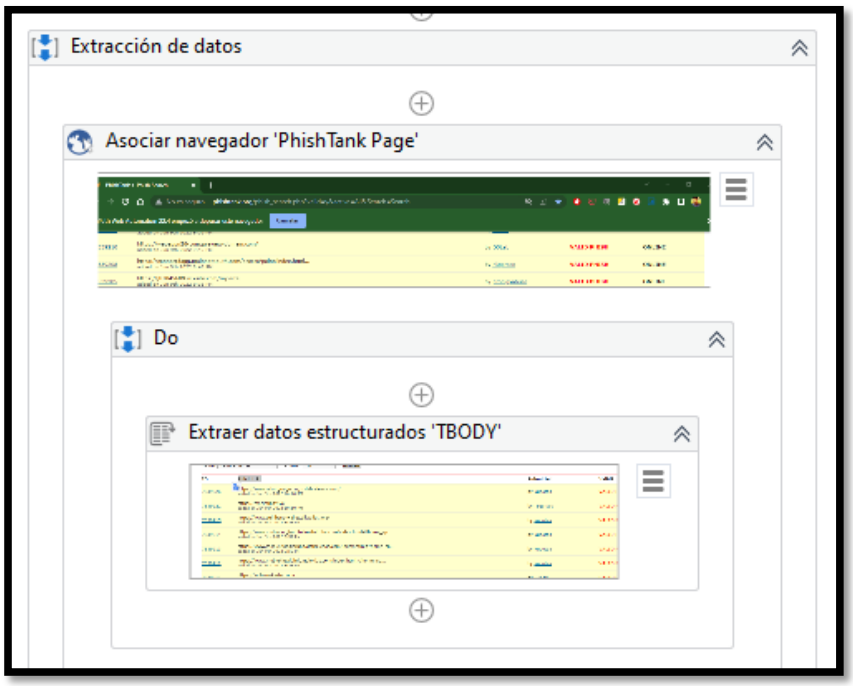
Figura 3: Evaluación mediante un algoritmo

Fuente: somosasistentes.com



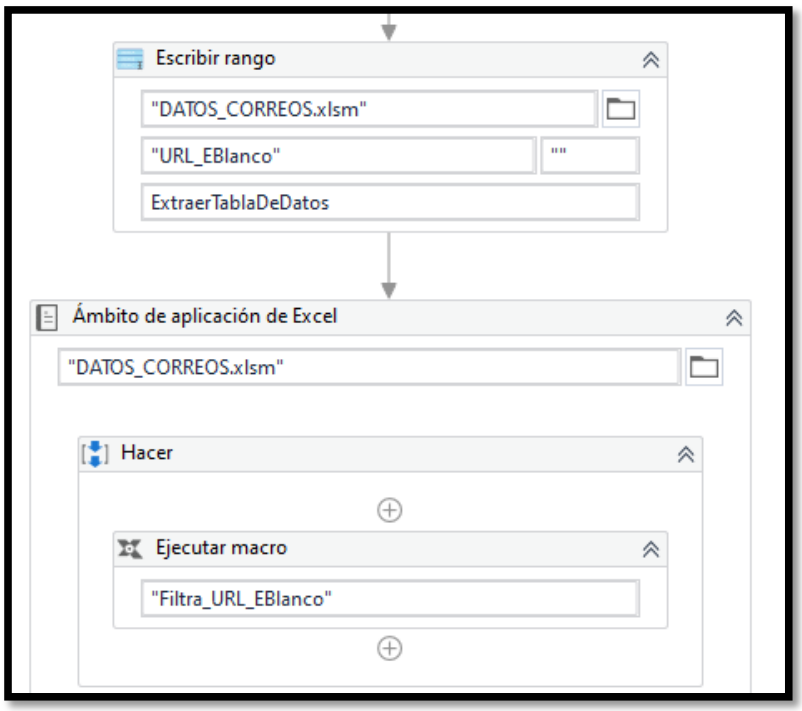
En la siguiente figura nos nuestra como una búsqueda utilizando web scraping con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 4: Búsqueda utilizando web scraping



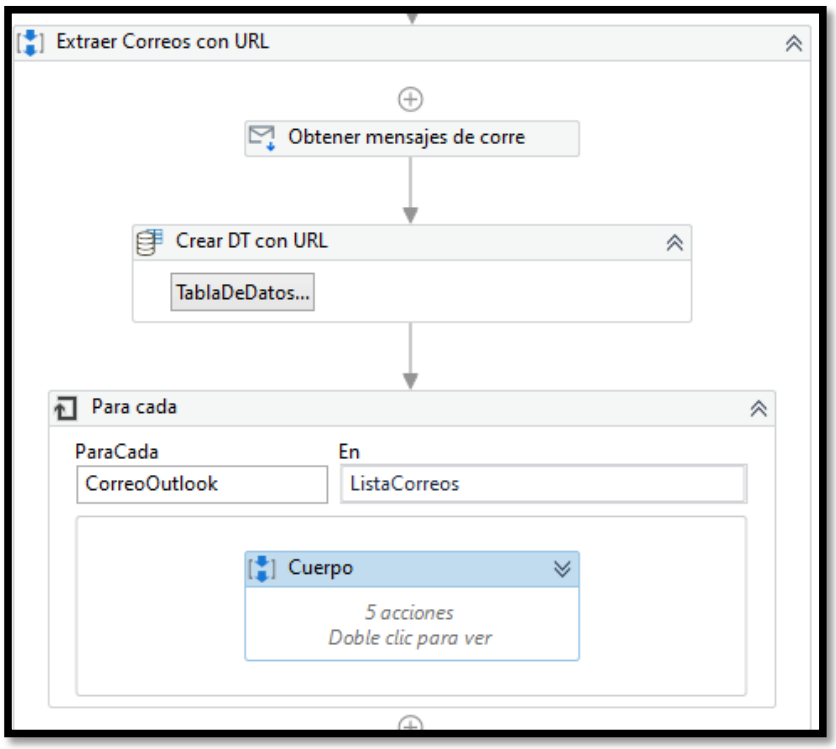
En la siguiente figura nos nuestra como se guarda la búsqueda en Excel de URL malicioso con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 5: Guardar en Excel de URL malicioso



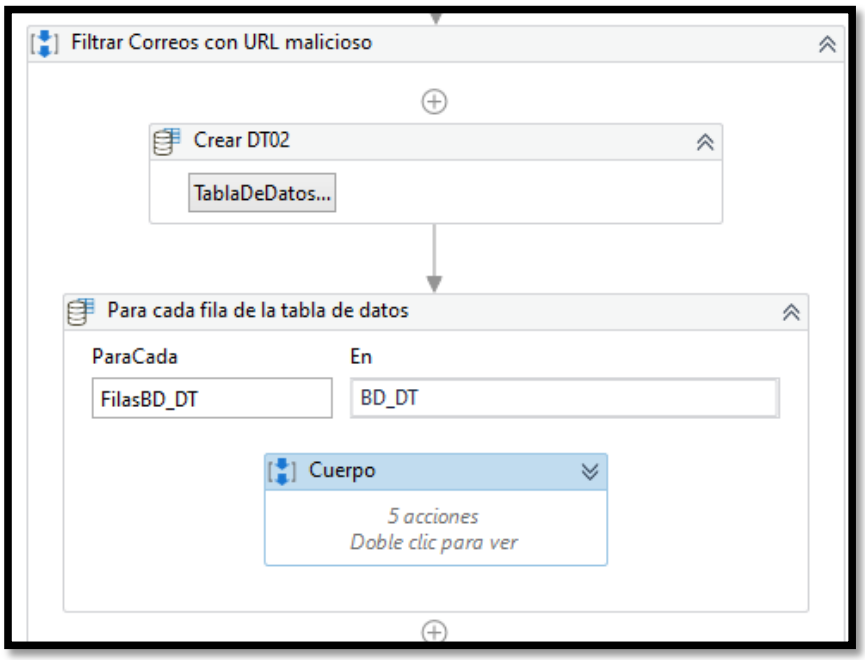
En la siguiente figura nos nuestra como una evaluación los correos que contenga URL con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 6: Evaluación los correos que contenga URL



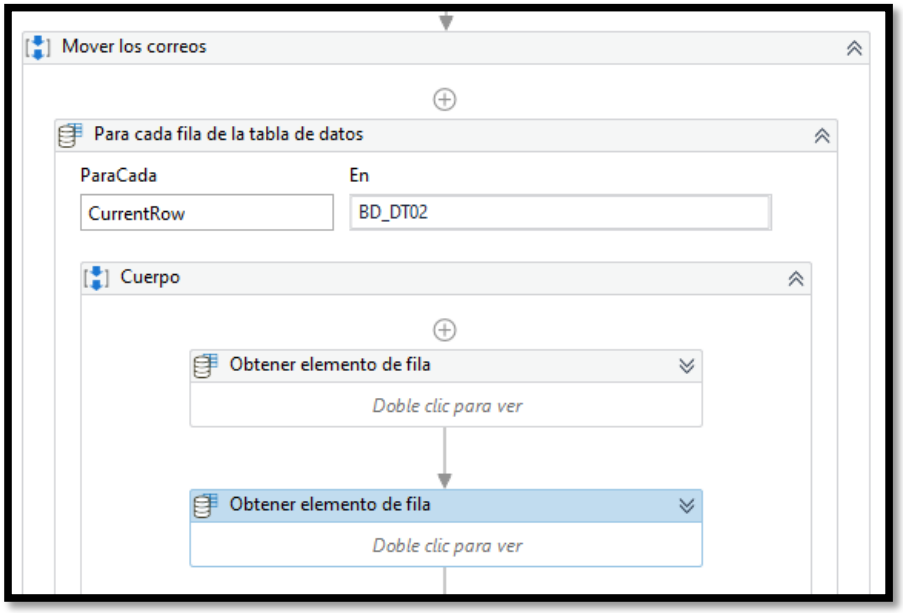
En la siguiente figura nos nuestra como una evaluación mediante la búsqueda de URL malicioso dentro de los correos con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 7: Búsqueda de URL malicioso



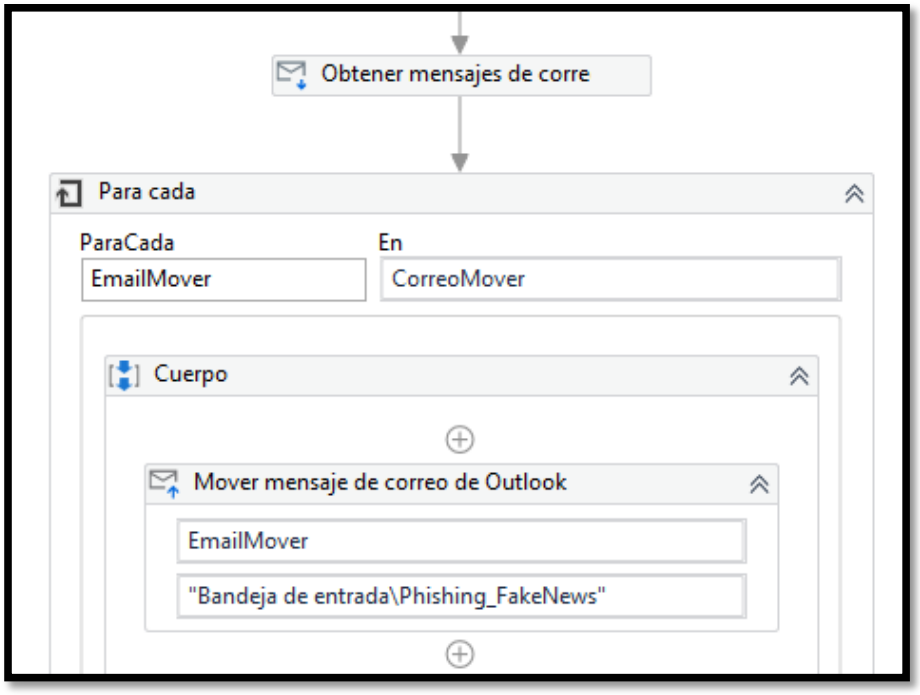
En la siguiente figura nos nuestra como la búsqueda de correos almacenados en la base de datos temporal con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 8: Correos almacenados en la base de datos



En la siguiente figura nos nuestra como se procede con mover los correos con URL malicioso dentro de los correos con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 9: Mover los correos con URL malicioso



Anexo 6: Algoritmos de base

En la siguiente figura nos nuestra como accede al buzón principal con la aplicación RPA para la clasificación de correos electrónicos como phishing y fake news

Figura 10: Algoritmo ELM

Fuente: Luo et al., 2021

Algorithm: ELM

Input：sample set $\{(x_i, y_i) \mid x_i \in R^n, y_i \in R^m, i = 1, 2, ..., N\}$, the number of hidden layer nodes L, activation function $g(x)$

Output：output weight β

1) Randomly generate the values for input weight w_i and hidden layer offset b_i;

2) Calculate hidden layer output matrix H;

3) Calculate output weight β: $\hat{\beta} = H^{+}Y$.

Figura 11: Algoritmo k-step CD-K

Fuente: Yi et al., 2018

Require: Visible Layer $V = \{v_1, ..., v_m\}$, Hidden Layer $H = \{h_1, ..., h_n\}$

Ensure: Gradient Approximation $\Delta\theta \longleftarrow \Delta w_{ij}, \Delta a_i, \Delta b_j$ for i in {1...n}, j in {1...m}

for i in {1...n}, j in {1...m} **do**
  Initialize $\Delta w_{ij} = \Delta a_i = \Delta b_j = 0$
end for
for Each v in V **do**
  $v^0 \longleftarrow v$
  for t in {0...$k-1$} **do**
    for i in {1...n} **do**
      Sample $h_i^t \sim p(h_i|v^t)$
    end for
    for j in {1...m} **do**
      Sample $v_j^t \sim p(v_j|h^t)$
    end for
  end for
end for
for i in {1...n}, j in {1...m} **do**
  $\Delta w_{ij} \longleftarrow \Delta w_{ij} + p(h_i|v^0)v_j^0 - p(h_i|v^k)v_j^k$
  $\Delta a_i \longleftarrow \Delta a_i + p(h_i|v^0) - p(h_i|v^k)$
  $\Delta b_j \longleftarrow \Delta b_j + v_j^0 - v_j^k$
end for

UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Declaratoria de Autenticidad del Asesor

Yo, ALFARO PAREDES EMIGDIO ANTONIO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - LIMA ESTE, asesor de Tesis titulada: "Aplicación RPA para la clasificación de correos electrónicos como phishing y fake news", cuyo autor es ICOCHEA RIVERA GEORGE EDUARD, constato que la investigación cumple con el índice de similitud establecido, y verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

LIMA, 23 de Julio del 2022

Apellidos y Nombres del Asesor:	**Firma**
ALFARO PAREDES EMIGDIO ANTONIO **DNI:** 10288238 **ORCID** 0000-0002-0309-9195	Firmado digitalmente por: EALFAROP el 26-07-2022 13:38:39

Código documento Trilce: TRI - 0363455

INVESTIGA UCV