Maestría en

CIBERSEGURIDAD

Trabajo previo a la obtención de título de
Magister en Ciberseguridad

AUTOR/ES:

Arévalo Altamirano Julio Daniel

Malave Yela Roberto Francisco

Robinson Casierra Ray Ricardo

Silva Naranjo Bryan Patricio

TUTORES:

Paulina Vizcaíno

Iván Reyes Chacón

TEMA:

Análisis comparativo del phishing tradicional y del phishing generado por IA con evaluación experimental de modelos de detección en entornos corporativos.

Quito - Ecuador

junio – 2026

REINVENTEMOS EL FUTURO

Resumen

El phishing sigue siendo una de las amenazas más frecuentes dentro de entornos empresariales que utilizan el correo electrónico como uno de sus principales medios de comunicación. En los últimos años, las herramientas basadas en inteligencia artificial (IA) han permitido generar correos cada vez más convincentes, con una redacción más natural, con menos errores ortográficos y un contexto más acertado, lo que dificulta su detección por parte de los usuarios y de algunos mecanismos tradicionales de seguridad. El presente trabajo tuvo como objetivo analizar las diferencias entre el phishing tradicional y el phishing generado por IA, así como evaluar el comportamiento de distintos modelos de detección aplicados al correo electrónico. Para lo cual, se realizó una revisión de literatura científica publicada entre 2020 y 2026 y posteriormente una fase experimental basada en técnicas de machine learning y deep learning. Se utilizaron datasets de Kaggle, Hugging Face, además de correos generados mediante inteligencia artificial y algunas muestras traducidas al español. Después del proceso de limpieza, normalización, eliminación de registros inválidos y balanceo de clases, se obtuvo un dataset final de 161.974 correos distribuidos de forma balanceada entre correos phishing y correos legítimos. Se evaluaron seis modelos de detección: Árbol de Decisión, Random Forest, Naive Bayes, LSTM, BiLSTM y Redes Neuronales Convolucionales (CNN). Para comparar los resultados se utilizaron las métricas accuracy, precision, recall y F1-score, dando mayor importancia al recall debido a la relevancia que tiene detectar correctamente los correos phishing dentro de un entorno corporativo. Los resultados obtenidos muestran que todos los modelos alcanzaron buenos resultados utilizando los parámetros de entrenamiento adecuados. Sin embargo, los modelos basados en deep learning obtuvieron un mejor desempeño, destacándose LSTM por su capacidad de identificar correos phishing y mantener un buen equilibrio entre la detección y los errores de clasificación. En general, los resultados permiten concluir que los modelos

basados en deep learning pueden ser una buena alternativa para reforzar los mecanismos de detección frente a campañas de phishing cada vez más sofisticadas.

Palabras Claves: phishing, inteligencia artificial, ciberseguridad, machine learning, deep learning, correo electrónico, dataset.

Abstract

Phishing continues to be one of the most common threats in corporate environments where email is one of the main communication channels. In recent years, Artificial Intelligence (AI) tools have made it possible to generate increasingly convincing emails with more natural writing, fewer spelling mistakes, and more accurate context, making them more difficult for both users and some traditional security mechanisms to detect. The objective of this study was to analyze the differences between traditional phishing and AI-generated phishing, as well as to evaluate the performance of different detection models applied to email classification. To achieve this, a literature review of scientific publications published between 2020 and 2026 was conducted, followed by an experimental phase based on machine learning and deep learning techniques. Public datasets from Kaggle and Hugging Face were used, together with AI-generated emails and a number of emails translated into Spanish. After the processes of data cleaning, normalization, removal of invalid records, and class balancing, a final dataset of 161,974 emails was obtained, equally distributed between phishing and legitimate emails. Six detection models were evaluated: Decision Tree, Random Forest, Naive Bayes, LSTM, BiLSTM, and Convolutional Neural Networks (CNN). The models were compared using the metrics accuracy, precision, recall, and F1-score, giving special importance to recall because of its relevance in identifying phishing emails within corporate environments. The results show that all models achieved good performance using the selected training parameters. However, the deep learning models obtained better overall results, with LSTM standing out for its ability to identify phishing emails while maintaining a good balance between detection capability and classification errors. Overall, the findings suggest that deep learning models can be a good alternative for strengthening phishing detection against increasingly sophisticated phishing campaigns.

Keywords: phishing, artificial intelligence, cybersecurity, machine learning, deep learning, email, dataset.