UIDE
Powered by
Arizona State University

eig Business School

Maestría en

Ciencia de Datos y Máquinas de Aprendizaje mención en Inteligencia Artificial

Trabajo de titulación previo a la obtención del título de Magíster en Ciencia de Datos y Máquinas de Aprendizaje con mención en Inteligencia Artificial

AUTORES:

Andrango Paillacho Stalyn Fernando.

Conrado Mena Josué Israel.

Espin Landivar Luis Sebastián.

Sulca Naranjo Diego Fernando.

TUTORES:

Alejandro Cortés Msc.

Iván Reyes Ch. Mgtr.

Implementación de un Sistema de Detección de Sitios Web de Phishing Basado en Análisis de URL, SSL y Contenido Utilizando Aprendizaje Profundo

Quito, diciembre 2024

Resumen

El Phishing es una de las amenazas que ha evolucionado a lo largo del tiempo, lo cual lo hace un problema constante debido a su adaptabilidad, y para el cual, los mecanismos tradicionales como filtros y listas negras no son suficientes; ante esto se hace necesario diseñar mecanismos inteligentes que permitan identificar estas amenazas conforme cambian.

Al utilizar modelos de aprendizaje profundo, podremos identificar estas amenazas con mayor precisión y reducir los falsos positivos, lo cual proporcionará junto con integraciones de seguridad una experiencia confiable para los usuarios.

Esta solución es escalable y adaptable, a futuro puede integrarse en diferentes plataformas y aplicaciones, que con cierto enfoque fomentará la importancia de la ciberseguridad e inteligencia artificial.

Implementaremos un modelo híbrido, que combinará una Red Neuronal Recurrente (RNN) para el análisis de datos textuales y una Red Neuronal Artificial (ANN) para el procesamiento de datos numéricos. El desarrollo presentará desde procesos de extracción, transformación y carga (ETL), hasta el entrenamiento y la evaluación del modelo seleccionado para garantizar su rendimiento y precisión.

Palabras Claves

Phishing, aprendizaje profundo, red neuronal recurrente (RNN), red neuronal artificial (ANN), ciberseguridad, inteligencia artificial (IA), escalabilidad, extracción, transformación y carga (ETL), precisión.

Abstract

Phishing is one of the threats that has evolved over time, which makes it a constant problem due to its adaptability, and for which traditional mechanisms such as filters and blacklists are not enough; therefore, it is necessary to design intelligent mechanisms to identify these threats as they change.

By using deep learning models, we will be able to identify these threats more accurately and reduce false positives, which together with security integrations will provide a reliable experience for users.

This solution is scalable and adaptable, going forward it can be integrated into different platforms and applications, which with some focus will further the importance of cybersecurity and artificial intelligence.

We will implement a hybrid model, which will combine a Recurrent Neural Network (RNN) for textual data analysis and an Artificial Neural Network (ANN) for numerical data processing. The development will present from extraction, transformation and loading (ETL) processes, to training and evaluation of the selected model to ensure its performance and accuracy.

Keywords

Phishing, deep learning (DL), recurrent neural network (RNN), artificial neural network (ANN), cybersecurity, artificial intelligence (AI), scalability, extraction, transformation and loading (etl), accuracy.