RESUMEN

El continuo surgimiento de nuevas tecnologías ha ocasionado el incremento en la delincuencia cibernética. Estos ciberataques se han convertido en amenazas graves, dando origen a nuevos malware o programas con código malicioso, en donde utilizan técnicas de Ingeniería Social con el fin de robar o destruir datos importantes. Por tal razón, los ciberdelincuentes aprovechan la ingenuidad de las personas para robar información confidencial, esto ha generado el incremento de fraudes durante los últimos cinco años. Frente a este escenario el presente proyecto se enfoca en el desarrollo de un modelo para la detección y mitigación de ataques de suplantación de identidad utilizando técnicas de Machine Learning. Cabe mencionar que el desarrollo se realizó en un ambiente controlado para garantizar la seguridad del entorno. Para la generación de correos infectados se extrajeron enlaces maliciosos de PhishTank. De modo que se realizó la extracción de las características de los correos para la fase de entrenamiento utilizando el algoritmo Naive Bayes. Luego se detectaron los correos infectados mediante el algoritmo de Árboles de Decisión con la finalidad de enviar a cuarentena los correos ilegítimos. Por último, se validó con los algoritmos de ML Random Forest, Regresión Logística y Clasificador Ficticio, con la finalidad de conocer el porcentaje de precisión en la detección de phishing de la solución propuesta en comparación con otros algoritmos de aprendizaje supervisado.

PALABRAS CLAVE:

- **INGENIERÍA SOCIAL**
- **SUPLANTACIÓN DE IDENTIDAD**
- **TÉCNICAS DE MACHINE LEARNING**
- **APRENDIZAJE SUPERVISADO**

ABSTRACT

The continuous emergence of new technologies has caused the increase in cybercrime. These cyber-attacks have become serious threats, giving rise to new malware or programs with malicious code, where they use social engineering techniques in order to steal or destroy important data. For this reason, cybercriminals take advantage of the ingenuity of people to steal confidential information; this has generated increased fraud during the last five years. Faced with this scenario, the present project focuses on the development of a model for the detection and mitigation of identity theft attacks using Machine Learning techniques. It is worth mentioning that the development was carried out in a controlled environment to guarantee the safety of the environment. For the generation of infected emails, malicious links were extracted from PhishTank. So the extraction of the mail characteristics for the training phase was carried out using the Naive Bayes algorithm. Then the infected emails were detected using the Decision Trees algorithm in order to quarantine the illegitimate emails. Finally, it was validated with the algorithms of ML Random Forest, Logistic Regression and Fictitious Classifier, with the purpose of knowing the percentage of accuracy in the phishing detection of the proposed solution in comparison with other supervised learning algorithms.

KEYWORDS:

- **SOCIAL ENGINEERING**
- **IDENTITY EXPLANATION**
- **MACHINE LEARNING TECHNIQUES**
- **SUPERVISED LEARNING**