Entrega 1

Responsables:

DAVID LONDONO LÓPEZ

ISAAC JIMÉNEZ FERNADEZ

Profesor:

Raul Ramos Pollan

Universidad de Antioquia

Introducción a la inteligencia artificial

2023/1

1. Planteamiento del problema

El problema que se está planteando es un problema de clasificación, que tiene como objetivo detectar y clasificar páginas con phishing. El término anti-phishing se refiere a las medidas preventivas para bloquear los ataques de phishing. El phishing es un delito cibernético en el que los atacantes se hacen pasar por entidades confiables o conocidas y contactan a las personas a través de diferentes medios, como correo electrónico, mensajes de texto o teléfono, con el fin de obtener información confidencial.

Por lo general, en un ataque de phishing por correo electrónico, el mensaje engañoso sugerirá que hay un problema con una factura, que ha habido actividad sospechosa en una cuenta o que el usuario debe iniciar sesión para verificar una cuenta o contraseña. Además, los atacantes pueden solicitar a los usuarios que ingresen información de la tarjeta de crédito, detalles bancarios y otros datos personales confidenciales.

Una vez que los atacantes recopilan esta información, pueden utilizarla para acceder a cuentas, robar datos e identidades, así como descargar malware en la computadora del usuario. Por lo tanto, la implementación de medidas de seguridad efectivas para prevenir los ataques de phishing es esencial para proteger la privacidad y seguridad en línea de los usuarios.

2.Dataset

El dataset seleccionado es Phishing Dataset for Machine Learning (https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning) Este conjunto de datos contiene 48 atributos extraídos de 5000 páginas web de phishing y 5000 páginas web legítimas, que se descargaron de enero a mayo de 2015 y de mayo a junio de 2017. Se emplea una técnica mejorada de extracción de funciones aprovechando el marco de automatización del navegador (es decir, Selenium WebDriver), que es más preciso y robusto en comparación con el enfoque de análisis basado en expresiones regulares.

Algunos de los atributos más significativos son:

- NumDots, variable continua numérica
- UrlLength, variable continua numérica
- NumDash, variable continua numérica
- NoHttps, variable categórica 0 y 1
- IpAddress, variable categórica 0 y 1
- RandomString, variable categórica 0 y 1
- HostnameLength, variable continua numérica
- PopUpWindow, variable categórica 0 y 1

3. Métricas

Para la evaluación del sistema se emplearán dos métricas de evaluación principales: el accuracy y el f1 score, ya que ambos se enfocan en la precisión.

Además de estas métricas técnicas, se tiene en cuenta la métrica de negocio, la fiabilidad de las predicciones para determinar si una página tiene phishing o no. Es crucial que estas predicciones sean confiables para que el navegador web pueda evitar que sus usuarios accedan a páginas maliciosas.

4. Desempeño

En un modelo de este tipo, se espera que la precisión de las predicciones sea alta, superando el 80%, además será importante evitar un gran número de falsos positivos.

En un ambiente productivo, el modelo sería utilizado como un filtro para prevenir que los usuarios accedan a páginas sospechosas y, de esta manera, garantizar la seguridad de los usuarios. Por lo tanto, es fundamental que el modelo pueda proporcionar predicciones confiables y precisas para lograr este objetivo.

Referencias

https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/regression/supporting-topics/basics/what-are-categorical-discrete-and-continuous-variables/

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html