Universidad del Valle de Guatemala
Facultad de Ingeniería
Departamento de Ciencias de la Computación
CC3094 - Security Data Science

Laboratorio #1 - Detección de Phishing

Andrea Amaya - 19357
Brandon Hernández - 19376
Guatemala, Ciudad de Guatemala 9 de febrero de 2023

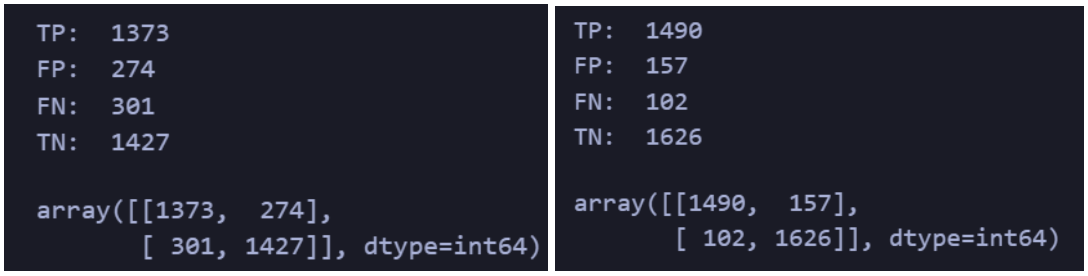
Discusión

- **¿Cuál es el impacto de clasificar un sitio legítimo como Phishing?**
  Esto puede causar que el usuario se vaya a otro sitio similar, que en este caso sí sea phishing, y meta sus credenciales dándole a un cracker un nuevo usuario al que puede vulnerar.

- **¿Cuál es el impacto de clasificar un sitio de Phishing como legítimo?**
  Esto puede causar que el usuario se quede en este sitio, ya que según el software esta página es legítima. Entonces el usuario ingresa sus credenciales y el cracker obtendrá nueva información.

- **En base a las respuestas anteriores, ¿Qué métrica elegiría para comparar modelos similares de clasificación de phishing?**
  Una matriz de confusión para visualizar de mejor manera cuántos falsos negativos se obtuvieron. El mejor de los casos es que todos los resultados fueran verdaderos positivos o verdaderos negativos.

- **¿Qué modelo es mejor para la clasificación de Phishing? Justifique**
  El modelo de machine learning con árboles de decisión tuvo un mejor resultado en comparación al modelo de machine learning k-NN.



k-NN / Árbol de decisión

El árbol de decisión tuvo aproximadamente 66% falsos negativos menos que el k-NN. Y, una precisión de 0.911 comparado al 0.828 de k-NN.

- **En base a las métricas obtenidas ¿es necesaria la intervención de una persona humana para tomar la decisión final?**
  Se recomienda que una persona verifique los falsos negativos para que los clasifique como verdaderos positivos o negativos. De esta forma no se están clasificando sitios de phishing cómo “no de phishing”, lo cual causaría que las personas caigan en robo de información debido a una mala clasificación del modelo.