1 Objetivos

- Conocer las características de sitios web destinados a Pishing
- Implementar un modelo de machine learning utilizando el algoritmo de árboles de decisión, para clasificar si un sitio es legítimo o no.

2 Preámbulo

Ingeniería Social

Consiste en la manipulación psicológica de una persona (comportamiento humano), con el fin de obtener información confidencial de ellos, que luego puede ser usada para comprometer un sistema.

Técnicas

- Baiting: convencer a la victima de revelar información, a cambio de una recompensa
- Impersonation: pretender ser alguien mas
- Dumpster diving: recolectar información (papeles con direcciones, correos, etc.) de los contenedores de basura
- Shoulder surfing: Espiar en las máquinas de otras personas, desde atrás, mientras las victimas tipean.
- Pishing: esta es la técnica más usada, ocurre cuando un atacante se enmascara como una entidad confiable, y engaña a una víctima para abrir un correo, mensaje instantáneo o un mensaje de texto.

3 Desarrollo

El laboratorio será desarrollado en parejas. Se debe entregar un enlace a un repositorio de Github con el reporte del perfil de datos, el código fuente de los modelos, el dataset de entrenamiento, validación y pruebas, y la explicación de las métricas de evaluación.

Parte 1 – Ingeniería de características

Exploración de datos

1. Cargue el dataset en un dataframe de pandas, muestre un ejemplo de cinco observaciones.
2. Muestre la cantidad de observaciones etiquetadas en la columna *status* como “legit” y como “pishing”. ¿Está balanceado el dataset?

Derivación de características

En base al artículo "Towards Benchmark Datasets for ML Based Wensite Phishing Detection: An Experimental Study", derivar las características basadas en el dominio: f1, f2, f4 – f20, f25, f26 y f27.

Para ello escriba las funciones necesarias y genere las nuevas columnas del dataset. Muestre un nuevo ejemplo de cinco observaciones donde se visualicen algunas de las columnas nuevas.

Preprocesamiento

Realice las modificaciones necesarias para convertir la variable categórica *status* a una variable binaria. Elimine la columna del dominio.

Visualización de resultados

Genere un reporte de perfil con la librería pandas_profiling. Analice el reporte y determine las columnas que son constantes, o que no tienen una varianza alta con la columna *status*. Almacene su reporte como una página html.

Selección de Características

En base al análisis del reporte, elimine las características repetidas o irrelevantes para la clasificación de un sitio de pishing. Verifique que no posee datos repetidos.

Parte 2 – Implementación del modelo

Separación de datos

- Datos de entrenamiento: 55%
- Datos de validación: 15%
- Datos de prueba: 30%
- Almacene cada dataset como un archivo .csv

Implementación

Utilice el algoritmo de árboles de decisión para entrenar el modelo. Muestre y explique los valores obtenidos de las siguientes métricas para los datos de validación y pruebas

- Matriz de confusión
- Precision
- Recall

- F1 Score

Responda las siguientes preguntas:

1. ¿Cuál es el impacto de clasificar un sitio legítimo como Pishing?
2. ¿Cuál es el impacto de clasificar un sitio de Pishing como legítimo?
3. En base a las respuestas anteriores, ¿Qué métrica elegiría para comparar modelos similares de clasificación de pishing?
4. ¿Es necesaria la intervención de una persona humana en la decisión final de clasificación?