1 Objetivos

- Conocer las características que mejor clasifican las URLs de pishing
- Implementar un modelo de ML para clasificar si un dominio es legítimo o es pishing.

2 Preámbulo

Phishing

Se basa en la ingeniería social (manipulación de emociones, aprovechamiento de atajos mentales y sesgos cognitivos) para engañar a las víctimas y lograr que estas den información (normalmente credenciales). Los atacantes envían mensajes haciéndose pasar por una entidad legitima a través de correos y SMS bajo diversos "motivos urgentes" que requieren que la persona tome acción inmediatamente, para lo cual incluyen un enlace que redirige al "sitio web" de la entidad.

Estos sitios son literalmente copias de los sitios legítimos que intentan imitar, en muchas ocasiones son muy difíciles de detectar. El usuario, temeroso de un evento negativo ingresa con sus credenciales, las cuales son robadas y utilizadas por los atacantes para acceder a los verdaderos sitios legítimos ocasionando pérdidas económicas (entre otros).

Sin embargo, los dominios web no pueden copiarse al 100%, aunque existen técnicas avanzadas que los hacen parecer similares al ojo humano. Sin embargo, las URLs de phishing poseen características que las diferencian de las URLs legítimas, y que un modelo de ML puede utilizar para detectar antes que un usuario "pique."

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 y la explicación de las métricas de evaluación. Se proporcionará un dataset con URLs legítimas y de pishing. El lenguaje de programación a utilizar será Python.

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 a los artículos propuestos de clasificación de phishing, responda las siguientes preguntas:

1. ¿Qué ventajas tiene el análisis de una URL contra el análisis de otros datos, cómo el tiempo de vida del dominio, o las características de la página Web?
2. ¿Qué características de una URL son más prometedoras para la detección de phishing?

En base a la respuesta anterior escriba al menos **quince** funciones basadas en los artículos, para derivar características que un modelo pueda utilizar y añada dichas características al dataset original.

Preprocesamiento

Realice las modificaciones necesarias para convertir la variable categórica *status* a una variable binaria. Elimine la columna del dominio. Realice el pre-procesamiento necesario.

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 phishing. Verifique que no posee observaciones repetidas.

Parte 2 – Implementación

Separación de datos

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

Implementación

Utilice dos algoritmos de Machine Learning para entrenar el modelo. Muestre y explique los valores obtenidos de las siguientes métricas para los datos de validación y pruebas, para cada modelo, en base al contexto del problema (detección de Pishing).

- Matriz de confusión
- Precision
- Recall
- Curva ROC
- AUC

Discusión

3. ¿Cuál es el impacto de clasificar un sitio legítimo como phishing?
4. ¿Cuál es el impacto de clasificar un sitio de phishing como legítimo?
5. En base a las respuestas anteriores, ¿Qué métrica elegiría para comparar modelos similares de clasificación de phishing?
6. ¿Qué modelo funcionó mejor para la clasificación de phishing? ¿Por qué?
7. Una empresa desea utilizar su mejor modelo, debido a que sus empleados sufren constantes ataques de phishing mediante e-mail. La empresa estima que, de un total de 50,000 emails, un 15% son phishing. ¿Qué cantidad de alarmas generaría su modelo? ¿Cuántas positivas y cuantas negativas? ¿Funciona el modelo para el BR propuesto? En caso negativo, ¿qué se podría hacer para reducir la cantidad de falsas alarmas?

Rúbrica

Aspecto	**Punteo (sobre 100 pts)**
Preguntas 1 - 2	10
Ingeniería de características	15
Implementación completa de los dos modelos (10 pts c/u)	20
Explicación de las métricas de rendimiento para cada modelo	15
Preguntas 3 - 7	40