

Conjunto de datos

Para esta actividad se utilizará un conjunto de datos sobre ciberseguridad, el mismo está diseñado para detectar intrusiones según el tráfico de la red y el comportamiento del usuario. El dataset está contenido en los archivos **srt_1.csv** y **str_2.csv** y almacena las siguientes variables:

- ID
- protocol - Protocolo de comunicación.
- browser_type - Navegador utilizado.
- login_attempts - Intentos de login.
- duration - Duración de la conexión (Segundos).
- packet_size - Tamaño de paquete (Bytes).
- encryption_used - Protocolo de encriptación.
- ip_reputation_score - Confiabilidad del IP (Una puntuación de 0 a 1, donde los valores más altos indican actividad sospechosa).
- failed_logins - Intentos fallidos de login.
- unusual_time_access - Horario inusual.
- attack - Detección de ataque (1 significa que se detectó un ataque, 0 significa actividad normal).

Práctica

Utilizando Python, sin recurrir a otros paquetes externos que no sean los vistos en el curso:

- 1) Realizar un análisis exploratorio de datos utilizando distintos tipos de funciones, gráficas y medidas estadísticas. Presentar los resultados más importantes.
- 2) Preprocesar el conjunto de datos para que pueda utilizarse en modelos de aprendizaje automático teniendo en cuenta:
 - Análisis de las variables.
 - Valores erróneos/faltantes.
 - Valores atípicos.
 - Correlación entre variables.
 - Procesamiento de variables categóricas.
 - Balance del conjunto de datos.
 - Normalización del conjunto de datos.
- 3) Entrenar y mostrar los resultados obtenidos con al menos dos modelos vistos en el curso, teniendo en cuenta:
 - División del conjunto de datos para entrenamiento y prueba.
 - Búsqueda de hiperparámetros para mejorar el rendimiento de los modelos.
 - Evaluación del modelo con las métricas convenientes.
- 4) Entrenar y mostrar los resultados obtenidos con al menos un modelo ensamblado visto en el curso, teniendo en cuenta:
 - División del conjunto de datos para entrenamiento y prueba.
 - Búsqueda de hiperparámetros para mejorar el rendimiento del modelo.
 - Evaluación del modelo con las métricas convenientes.

Entrega

- 1) Un archivo .ipynb conteniendo la solución a los incisos 1, 2, 3 y 4.
- 2) Un informe que documente y justifique los distintos pasos realizados, herramientas y técnicas utilizadas.
- 3) El trabajo se defenderá en un coloquio en fecha y hora a publicar.