Traccia 1: Realizzazione di un sistema di malware detection

Obiettivo: Realizzare un sistema di malware detection robusto utilizzando metodologie di machine learning e deep learning

1. Selezionare tre metodologie tra quelle viste durante il corso sia di machine learning che di deep-learning (almeno una deve essere basata su ML)
2. Addestrare e validare i classificatori sul dataset 1
3. Valutare l'accuratezza dei metodi scelti sul dataset 1
4. Valutare la capacità di generalizzazione dei metodi scelti sul dataset 2
5. Generare un dataset di campioni offuscati, partendo da quelli presenti nel dataset 2, tramite l'attacco GAMMA applicato sulla rete MalCONV messa a disposizione dalla libreria SECML
6. Valutare la robustezza dei classificatori selezionati sui campioni offuscati con diverse intensità dell'attacco
7. Produrre un documento di dettaglio della sperimentazione effettuata, descrivendo le procedure di estrazione delle feature, di addestramento e validazione e commentando i risultati sperimentali

Nota: per i gruppi da 3 persone i punti 5 e 6 diventano.

- Generare un dataset di campioni offuscati, partendo da quelli del dataset 2, aggiungendo i padding alla fine del file binario composto da byte generati con valori casuali.
- Valutare la robustezza dei classificatori selezionati sui campioni offuscati con diverse intensità dell'attacco, dove l'intensità è rappresentata dal rapporto il numero di byte aggiunti e la dimensione originale del binario.

Dataset:

- Dataset 1:
  - https://drive.google.com/file/d/1xBxl3ZOVj-MBCOIby5OP2k1i7iDhHDUU/view
- Dataset 2:
  - https://drive.google.com/open?id=1dB0rqJJweTPzdc4RENoJewtGtEQdWXzj

Traccia 2: Valutazione della sicurezza di un sistema di speaker recognition

Obiettivo: Realizzare e valutare le performance di un sistema di controllo accessi basato su speaker recognition.

1. Estrarre le feature con la rete fornita a lezione da tutti i campioni presenti nel dataset distribuito per il progetto.
2. Addestrare e validare con tali features due classificatori binari a scelta (es. MLP, KNN), la cui classe positiva è costituita dalle voci di una classe target a scelta (l'utente autorizzato), mentre la classe negativa contiene le voci di tutte le altre persone.
3. Testare i classificatori su un dataset di test costituito da campioni non corrotti
4. La divisione del dataset fornito in training, validation e test set, così come la scelta della classe target (l'utente autorizzato), è a discrezione dei gruppi
5. Generare degli adversarial examples utilizzando la libreria ART e uno dei due classificatori.
6. Valutare e discutere l'effetto degli attacchi sul classificatore utilizzato come white box.
7. Verificare se i campioni avversari hanno effetto anche sull'altro classificatore (trasferibilità).
8. Implementare e valutare le performance di almeno un meccanismo di difesa a scelta e specificare come certificare la difesa rispetto agli attacchi considerati.
9. Produrre un documento di dettaglio della sperimentazione effettuata, descrivendo le procedure di addestramento e validazione, la generazione degli attacchi, la generazione e la valutazione della difesa e commentando i risultati sperimentali.

Dataset:

- Voci: https://drive.google.com/file/d/1ViQ1tMzPwA6fA3S3_jG84uuKd0OvflZy/view?usp=sharing
- Annotazioni identità: https://drive.google.com/file/d/1ZXR1n_Ei-uuuRheuOkx3TaMhLtPDCazE/view?usp=sharing