Data Analytics for Cyber Security Lab2:

Proposing an ML Architecture and Standards for Generating IDS Scripts

1. Objective

This lab will guide students in proposing an architecture for a machine learning (ML) model that learns from data and establishing standards to generate Intrusion Detection System (IDS) scripts. Students will also implement a Python script for analyzing data, understanding regulations, and generating IDS scripts using a large language model (LLM).

2. Lab Outline

1. **Understanding ML Architectures**
   - Key components of an ML system.
   - Challenges in designing architectures for IDS applications.
2. **Standards for IDS Script Generation**
   - Overview of IDS scripts (e.g., Snort, Suricata).
   - Regulatory compliance and best practices.
3. **Hands-on Activities**
   - Drawing the architecture.
   - Writing a Python script for data analysis.
   - Using LLMs for IDS script generation.
4. **Assessment Questions**

1. Understanding ML Architectures

Key Components:

- **Data Collection:** Sources of data (e.g., network traffic logs).
- **Data Preprocessing:** Cleaning, normalization, and feature extraction.
- **Model Training:** Algorithms suitable for IDS (e.g., Random Forest, Neural Networks).
- **Model Evaluation:** Metrics (e.g., accuracy, recall, precision).
- **Deployment:** Integration with IDS systems.

Design Considerations:

- Scalability to handle large datasets.
- Real-time inference capability.
- Interpretability of results.

2. Standards for IDS Script Generation

Key Aspects:

- **Syntax and Structure:** Follow **Snort** or **Suricata** conventions.
- **Regulatory Compliance:** Ensure alignment with local and international cybersecurity standards (e.g., GDPR, NIST).
- **Performance Standards:** Minimize false positives and ensure scripts are optimized for speed.

3. Hands-on Activities

Activity 1: Drawing the Architecture

Task:

- Draw an architecture for an ML-based IDS system.
- Include components for data collection, preprocessing, training, evaluation, and deployment.

Questions:

1. Identify and describe the role of each component in your architecture.
2. How would you ensure the architecture is scalable and secure?

Activity 2: Python Script for Data Analysis and IDS Script Generation

Script Objective: Analyze network data and use an LLM to generate IDS rules.

Instructions:

1. Write a Python script that:
    - Loads a dataset of network logs.
    - Extracts suspicious patterns using statistical methods.
    - Passes patterns to an LLM for IDS rule generation.
2. Ensure the script adheres to regulatory standards

4. Assessment Questions

1. **Architecture Design:**
    - Sketch the architecture for the ML-based IDS system.

    - Explain how your design ensures low latency and high accuracy.

2. **Python Script:**
    - Identify key components in the provided Python script.
    - Modify the script to include additional preprocessing steps, such as normalization.
3. **Regulatory Compliance :**
    - What are the key cybersecurity regulations relevant to IDS systems in your region?
    - How can you ensure the generated IDS scripts comply with these regulations?

Deliverables

- A drawn architecture with explanations.
- Python script implementation.
- Answered assessment questions.