Bayesian Risk Assessment Using Cybersecurity Data

Objective:

The goal of this project is to apply **Bayesian statistics** to real-world cybersecurity data in order to assess the risk of successful cyberattacks on various assets in a network. You will use pre-processed data provided to build a **Bayesian Network** and compute the probabilistic risk of exploitation for different assets. The ultimate goal is to provide a ranked list of assets based on their risk levels, along with actionable insights for mitigation.

Problem Overview:

You are provided with four pre-processed sample datasets that describe the relationships between:

- **Assets**: Critical infrastructure like servers, databases, and applications.
- **Vulnerabilities**: Security weaknesses (mapped to CVE IDs) that could be exploited by attackers.
- **Attack Vectors**: Methods used by threat actors to exploit vulnerabilities (e.g., phishing, SQL injection).
- **Threat Actors**: Types of attackers (e.g., insider threats, external hackers) that may target specific assets.

Your task is to:

1. **Construct a Bayesian Network** based on the given data.
2. **Calculate the posterior probabilities** of successful exploitation for each asset by integrating prior attack probabilities, conditional probabilities of attack success, and relationships between assets, vulnerabilities, and attack vectors.
3. **Rank assets based on risk** and provide insights into which assets are most vulnerable.
4. **Recommend mitigation strategies** based on your risk findings.

Datasets for Bayesian Risk Assessment

Dataset 1: Asset-Vulnerability Mapping (25 sample entries)

This dataset contains the mapping between organizational assets and known vulnerabilities, along with the associated **CVSS score** (severity) and the **probability of exploitation**:

Asset	**Vulnerability**	**CVSS Score**	**Exploit Probability**
Server-01	CVE-2023-12345	9.0	0.75
Workstation-01	CVE-2023-54321	7.5	0.55
Database-01	CVE-2023-67890	8.5	0.70
WebApp-01	CVE-2023-09876	9.1	0.68
IoTDevice-01	CVE-2023-67891	6.8	0.45
Server-02	CVE-2023-87654	8.9	0.80
Server-03	CVE-2023-54322	7.2	0.55
Database-02	CVE-2023-98765	8.7	0.72
Workstation-02	CVE-2023-54321	7.5	0.58
WebApp-02	CVE-2023-23456	8.0	0.60
Database-03	CVE-2023-76543	8.3	0.64
IoTDevice-02	CVE-2023-34567	7.0	0.50
Server-04	CVE-2023-45678	9.0	0.75
Workstation-03	CVE-2023-98765	8.7	0.68
IoTDevice-03	CVE-2023-23456	8.1	0.63
Server-05	CVE-2023-23456	8.1	0.67
Database-04	CVE-2023-87654	8.9	0.70
WebApp-03	CVE-2023-87653	8.5	0.65
IoTDevice-04	CVE-2023-67890	8.5	0.60
Workstation-04	CVE-2023-45678	9.0	0.70
Server-06	CVE-2023-76543	8.3	0.75
WebApp-04	CVE-2023-67891	6.8	0.62
Database-05	CVE-2023-98765	8.7	0.71
IoTDevice-05	CVE-2023-54321	7.5	0.50

WebApp-05	CVE-2023-12345	9.0	0.65

Dataset 2: Attack-Vulnerability Mapping (25 entries)

This dataset shows the relationship between known vulnerabilities and potential attack vectors. Each row indicates the **success probability** of an attack vector exploiting a particular vulnerability:

Vulnerability	**Attack Vector**	**Success Probability**
CVE-2023-12345	Remote Code Execution	0.80
CVE-2023-54321	Phishing	0.60
CVE-2023-67890	SQL Injection	0.85
CVE-2023-54321	Phishing	0.60
CVE-2023-98765	SQL Injection	0.75
CVE-2023-09876	Phishing	0.55
CVE-2023-67891	Remote Code Execution	0.65
CVE-2023-87654	SQL Injection	0.80
CVE-2023-54322	RCE	0.75
CVE-2023-76543	Phishing	0.50
CVE-2023-23456	SQL Injection	0.70
CVE-2023-34567	RCE	0.60
CVE-2023-45678	SQL Injection	0.85
CVE-2023-98765	RCE	0.65
CVE-2023-76543	RCE	0.65
CVE-2023-87653	Phishing	0.50
CVE-2023-67891	SQL Injection	0.75
CVE-2023-12345	SQL Injection	0.75
CVE-2023-54322	SQL Injection	0.85
CVE-2023-87654	RCE	0.60

CVE-2023-09876	SQL Injection	0.80
CVE-2023-54321	RCE	0.70
CVE-2023-76543	SQL Injection	0.70
CVE-2023-34567	Phishing	0.55
CVE-2023-12345	SQL Injection	0.80

Dataset 3: Threat Actor-Asset Targeting (25 entries)

This dataset shows which threat actors are more likely to target certain assets and includes the **probability** of an attack by the given actor on a specific asset:

Threat Actor	**Asset**	**Target Probability**
Insider	Server-01	0.40
External Hacker	WebApp-01	0.70
Insider	Database-01	0.30
External Hacker	Workstation-01	0.50
Insider	Server-02	0.45
Insider	WebApp-02	0.50
Insider	IoTDevice-01	0.20
External Hacker	Workstation-02	0.60
External Hacker	Server-03	0.55
External Hacker	IoTDevice-02	0.45
Insider	Server-04	0.35
External Hacker	WebApp-03	0.75
Insider	IoTDevice-03	0.25
External Hacker	Database-02	0.65
Insider	WebApp-04	0.40
Insider	Workstation-03	0.50
External Hacker	Server-05	0.70

External Hacker	Workstation-04	0.60
Insider	Database-03	0.35
External Hacker	IoTDevice-04	0.55
Insider	Server-06	0.45
External Hacker	WebApp-05	0.60
External Hacker	Database-05	0.65
Insider	IoTDevice-05	0.40
External Hacker	Server-05	0.70

Dataset 4: Prior Attack Success Rate (25 entries)

This dataset shows the prior probabilities of successful attacks by threat actors using various attack vectors:

Threat Actor	**Attack Vector**	**Success Rate**
Insider	Phishing	0.35
External Hacker	SQL Injection	0.60
External Hacker	Remote Code Execution	0.65
Insider	Remote Code Execution	0.50
Insider	SQL Injection	0.40
External Hacker	Phishing	0.55
External Hacker	RCE	0.65
Insider	SQL Injection	0.35
External Hacker	RCE	0.70
Insider	SQL Injection	0.50
Insider	Phishing	0.45
External Hacker	Phishing	0.60
Insider	RCE	0.55
External Hacker	SQL Injection	0.75

Insider	SQL Injection	0.50
External Hacker	RCE	0.80
Insider	Phishing	0.40
External Hacker	SQL Injection	0.85
Insider	RCE	0.60
External Hacker	Phishing	0.65
Insider	SQL Injection	0.45
External Hacker	SQL Injection	0.75
External Hacker	Phishing	0.70
Insider	Phishing	0.50
External Hacker	RCE	0.80

Deliverables:

1. **Submission:**
    - Push all your code and relevant files to your favourite public git repository.
    - Ensure your commit history shows incremental progress.
    - Share the repository link with us.
2. **Brief Report: A brief report (1-2 pages) that:**
    - Explains your approach and methodology including Bayesian Network Diagram (A diagram showing the relationships between threat actors, attack vectors, vulnerabilities, and assets).
    - Posterior Probabilities and Asset Ranking: A list of assets ranked by their overall risk, with the posterior probabilities of successful attack for each asset.

Evaluation Criteria:

- **Correctness** of the Bayesian Network structure and methodology.
- **Accuracy** in computing posterior probabilities and ranking assets.
- **Clarity** and depth of the final report, including insights and findings.
- **Code quality** and clear use of version control (commit messages showing incremental progress).

Suggested tools (but, feel free to use whatever you see fit) :

- You can use either Python (e.g., pgmpy, PyMC3) or R (e.g., bnlearn, gRain) for building and analyzing the Bayesian Network.

- Visualization tools like matplotlib, Graphviz, or ggplot2 can be used for network diagrams.
- ChatGPT or any other AI tool can be used to generate code as long as you know exactly what it is doing and there is proof of it in your git commits.