**TASK NO 4**

**APPLYING THE DATA SCIENCE LIFECYCLE TO A REAL-WORLD APPLICATION(CYBERSECURITY-THREAT DETECTION)**

**1. PROBLEM DEFINITION (UNDERSTANDING THE THREATS)**

* Identify cybersecurity threats like malware, phishing, DDoS attacks, and unauthorized access.
* Define the goal: Detect anomalies and prevent cyberattacks in real-time.

**2. DATA COLLECTION (GATHERING SECURITY DATA)**

* Collect data from firewalls, intrusion detection systems (IDS), log files, network traffic, system logs, and endpoint security software.
* Include various formats: structured (logs, packet data) and unstructured (emails, messages).

**3. DATA PREPROCESSING & CLEANING**

* Remove irrelevant data and filter out noise from system logs.
* Convert raw network traffic data into a structured format for analysis.
* Handle missing values and normalize data for machine learning models.

**4. EXPLORATORY DATA ANALYSIS (EDA) & FEATURE ENGINEERING**

* Identify normal vs. abnormal patterns in user activity and network behavior.
* Extract useful features such as IP addresses, login times, file access patterns, and request frequencies.
* Visualize attack trends using graphs and anomaly detection plots.

**5. MODEL BUILDING & MACHINE LEARNING**

* Train supervised ML models (Random Forest, SVM, Neural Networks) for known attack detection.
* Use unsupervised ML models (Isolation Forest, Autoencoders) for anomaly detection.
* Apply deep learning (LSTMs, CNNs) for advanced attack pattern recognition.

**6. MODEL EVALUATION & OPTIMIZATION**

* Evaluate model performance using precision, recall, F1-score, and confusion matrix.
* Fine-tune hyperparameters to improve threat detection accuracy.
* Reduce false positives and false negatives to enhance reliability.

**7. DEPLOYMENT & REAL-TIME MONITORING**

* Deploy the trained model in SIEM (Security Information and Event Management) systems.
* Continuously monitor network activity and detect anomalies in real-time.
* Automate threat response by integrating with firewalls and incident response systems.

**8. MODEL MAINTENANCE & IMPROVEMENT**

* Periodically retrain models with new attack data to keep up with evolving cyber threats.
* Use reinforcement learning for adaptive cybersecurity threat detection.