**🔐 Project Title: Enhanced Intrusion Detection using CICIDS 2017 with Anomaly Detection, XAI & Visualization**

**🎯 Objective:**

To detect and analyze network intrusions using the CICIDS 2017 dataset by combining **supervised** and **unsupervised** learning approaches along with **visual and explainable AI techniques** to build a robust and insightful cybersecurity solution.

**📊 Dataset Used:**

* **CICIDS 2017** – A real-world, labeled dataset that includes both benign and attack traffic (e.g., DDoS, Port Scans, Brute Force, Botnet attacks).
* Includes 80+ network features per record like flow duration, packet size, protocol, etc.

**🔧 Key Steps and Techniques:**

**✅ 1. Preprocessing and Feature Selection:**

* Handled missing values, dropped non-numeric columns.
* Normalized the dataset using **StandardScaler**.

**✅ 2. Exploratory Data Analysis (EDA):**

* Visualized class distribution, feature correlations.
* Identified imbalance in attack vs normal traffic.

**✅ 3. Unsupervised Anomaly Detection (Unique):**

* Used **Isolation Forest** to detect anomalies without labels.
* Applied **UMAP** for dimensionality reduction and visualized the anomaly clusters clearly in 2D.

**✅ 4. Clustering (K-Means) of Attacks (Unique):**

* Discovered hidden patterns in traffic using **K-Means** clustering.
* Plotted using UMAP to reveal cluster-based groupings of network behavior.

**✅ 5. Explainable AI with SHAP (Unique):**

* Explained the decision-making process of the ML model using **SHAP values**.
* Displayed which features contributed most to identifying a record as an attack.

**✅ 6. Geo-Visualization of Attack Origins (Creative):**

* Simulated a **world map view** of attack origins using **Folium** (interactive HTML map).
* Useful for showing how global threats could look in a real deployment.

**✅ 7. Classification (Optional Section):**

* Used traditional models like **Random Forest**, **XGBoost**, or **SVM** for classifying network traffic.
* Achieved high accuracy in detecting known attack types.

**🧠 Why This Project is Unique:**

* Goes **beyond classification** to include:
  + Anomaly detection without labels,
  + Visual understanding through UMAP,
  + Interpretability via SHAP,
  + A geographic map to mimic real-world threat scenarios,
  + AutoML-style unsupervised clustering for cyber threat grouping.

**📌 Technologies Used:**

* Python, Pandas, Scikit-learn, SHAP, UMAP, Matplotlib, Seaborn, Folium

**🌟 Conclusion:**

This project not only detects intrusions with machine learning but also provides **explainable and visual insights** into network security, making it highly practical and presentable for real-world cyber defense systems.