**Experimental Setup**

The experiment is designed to develop and evaluate an **intrusion detection system (IDS)** by integrating **machine learning techniques** with **Open-Source Intelligence (OSINT)** data. The notebook (KDD99\_Intrusion\_Detection\_Final\_with\_OSINT\_main.ipynb) implements multiple supervised learning models, including **Logistic Regression**, **Decision Tree**, and **Random Forest**, to classify network traffic as *normal* or *attack*.

The workflow follows a structured pipeline:

1. **Data Preprocessing** – The raw dataset is cleaned, normalized, and encoded to prepare it for model training. Irrelevant attributes are removed, and categorical features are transformed into numerical form.
2. **Feature Engineering** – Key network features are extracted and combined with OSINT-based indicators, such as known CVE identifiers and threat intelligence attributes obtained from **ThreatMiner** and **AlienVault (OTX)**.
3. **Model Training and Validation** – The dataset is split into training and testing subsets (commonly 80/20). Each model is trained to detect intrusion patterns, and its performance is evaluated using **accuracy, precision, recall, and F1-score** metrics.
4. **Model Comparison** – Final results are compared across classifiers to determine the optimal model for IDS deployment. The Random Forest classifier typically shows superior results with accuracy above 99%.

**Data Sources**

1. **Primary Dataset:**
   * **File:** dataset.csv
   * **Description:** Derived from the **KDD99 intrusion detection dataset**, a widely used benchmark in cybersecurity research for anomaly and attack detection.
   * **Content:** Contains both *normal* and *attack* network traffic records with multiple attributes representing network connection features (e.g., duration, protocol type, service, flag, src\_bytes, dst\_bytes, etc.).
   * **Label:** The target variable classifies each record as either “attack” or “normal.”
2. **External OSINT Sources:**
   * **ThreatMiner** – Provides structured information about **CVE vulnerabilities**, exploit references, and malware associations.
   * **AlienVault OTX (Open Threat Exchange)** – Supplies **real-time threat indicators (IOCs)** such as malicious IPs, URLs, and exploit activity data.
   * **Integration:** These sources are programmatically queried and linked to corresponding attack features within the dataset, enriching the training data with real-world threat context.

**Purpose**

The integrated setup aims to enhance the traditional machine learning-based IDS by embedding **external threat intelligence** (OSINT) to improve adaptability, contextual awareness, and detection accuracy, particularly against emerging or zero-day attacks.