

## **Models Used for Cyber Threat Analysis**

Cyber threat analysis involves extracting meaningful insights from large volumes of heterogeneous security data such as attack logs, vulnerability feeds, and incident records. Various analytical and machine learning models are commonly used in cybersecurity research. The following models were studied and evaluated for applicability to the proposed **Interactive Cyber Threat Visualization Dashboard**.

### **1. Descriptive Statistical Analysis Model**

This model summarizes cybersecurity data using basic statistical measures such as frequency, count, percentage, and distribution. It helps convert raw incident logs into structured insights.

In cybersecurity, descriptive analysis is used to measure how often attacks occur, which attack types are most common, and how severity levels are distributed. In this project, this model supports bar charts, pie charts, and summary metrics that provide a high-level overview of the threat landscape.

#### **How it works:**

Uses statistical measures such as count, mean, frequency, and distribution to summarize security events.

#### **Uses in Cybersecurity:**

- Attack frequency analysis
- Severity distribution
- Trend visualization

#### **Accuracy:**

High for historical and aggregated data (depends on data quality)

#### **Relevance to Project:**

Forms the foundation for time-series charts, bar graphs, and severity analysis in the dashboard.

### **2. Time-Series Analysis Model**

Time-series analysis studies security events over time to identify patterns such as trends, spikes, and seasonal behavior. Each cyber incident is analyzed using its timestamp.

This model is important for detecting periods of increased attack activity and understanding how threats evolve. In the dashboard, time-series analysis enables line graphs that show daily, weekly, or monthly attack trends, helping analysts identify abnormal surges in cyber incidents.

#### **How it works:**

Analyzes data points indexed in time order to identify trends, spikes, and seasonal patterns.

#### **Uses in Cybersecurity:**

- Detecting attack surges
- Monitoring periodic attack behavior

- Identifying abnormal activity over time

**Accuracy:**

High for trend detection in structured temporal data

**Relevance to Project:**

Directly supports **trend and anomaly detection** outcomes using line charts and time filters.

### 3. Anomaly Detection Model (Statistical / Threshold-based)

Anomaly detection models identify unusual behavior by comparing current values against historical patterns or predefined thresholds.

In cybersecurity, this model helps detect sudden spikes in attack frequency or unexpected changes in severity levels. In the proposed system, anomalies are visually highlighted rather than automatically classified, allowing analysts to investigate suspicious periods interactively.

**How it works:**

Identifies deviations from normal behavior using statistical thresholds.

**Uses in Cybersecurity:**

- Identifying sudden spikes in attack frequency
- Detecting unusual severity levels

**Accuracy:**

Moderate to high for rule-based anomaly identification

**Relevance to Project:**

Used to visually highlight abnormal periods in attack timelines.

### 4. K-Means Clustering

K-Means clustering groups similar data points into clusters based on feature similarity such as attack type, severity, or target system.

This model is useful for exploratory cybersecurity analysis, where similar attack behaviors are grouped together. Although not a core component of the dashboard, clustering can assist in understanding relationships between attack categories during initial data analysis.

**How it works:**

Groups data points into clusters based on similarity.

**Uses in Cybersecurity:**

- Grouping attack types
- Clustering similar vulnerability patterns

**Accuracy:**

Moderate (depends on feature selection and K value)

**Relevance to Project:**

Useful for exploratory analysis but not essential for core visualization objectives.

## 5. Decision Tree Model

The Decision Tree model classifies data using a set of rule-based decisions formed in a tree structure.

In cybersecurity, it is commonly used for intrusion detection and attack classification. While decision trees provide clear decision rules, in this project they serve mainly as a reference model rather than a primary implementation, since the dashboard emphasizes visualization over automated classification.

**How it works:**

Uses rule-based splitting to classify data based on feature conditions.

**Uses in Cybersecurity:**

- Attack classification
- Rule-based threat categorization

**Accuracy:**

Moderate to high for structured labeled data

**Relevance to Project:**

Useful for analysis but not required for visualization-driven insights.

## 6. Random Forest Model

Random Forest is an ensemble learning model that combines multiple decision trees to improve accuracy and robustness.

This model is effective for predicting attack categories and identifying high-risk events. However, due to its computational complexity and lower interpretability, it is not directly integrated into the visualization dashboard, which prioritizes clarity and analyst-driven insights.

**How it works:**

Ensemble of decision trees that improves classification accuracy.

**Uses in Cybersecurity:**

- Threat classification
- Risk prediction

**Accuracy:**

High (often above 90% for labeled datasets)

**Relevance to Project:**

Better suited for prediction systems rather than visualization dashboards.

## **7. Support Vector Machine (SVM)**

Support Vector Machine separates data into classes by finding an optimal boundary between attack and normal behavior.

SVM is widely used in intrusion detection systems due to its strong classification performance. In this project, SVM is reviewed as part of the model study but is not implemented, as the focus remains on descriptive and visual analysis rather than classification.

### **How it works:**

Separates data using optimal hyperplanes.

### **Uses in Cybersecurity:**

- Intrusion detection
- Malware classification

### **Accuracy:**

High for binary classification problems

### **Relevance to Project:**

Not directly aligned with visualization-centric project goals.

## **8. Hierarchical Analysis Model**

Hierarchical analysis organizes cybersecurity data into parent-child relationships, such as attack categories and sub-techniques.

This model is particularly useful for mapping vulnerabilities to frameworks like MITRE ATT&CK. In the dashboard, hierarchical models enable treemap and sunburst visualizations that help analysts prioritize the most frequently targeted systems and techniques.

### **How it works:**

Organizes data into parent-child relationships.

### **Uses in Cybersecurity:**

- Vulnerability prioritization
- MITRE ATT&CK technique mapping

### **Accuracy:**

High for structured hierarchical data

### **Relevance to Project:**

Crucial for treemap and sunburst visualizations used in vulnerability analysis.

## **9. Geospatial Analysis Model**

Geospatial analysis uses geographical coordinates to visualize where cyber attacks originate and which regions are most targeted.

In cybersecurity, this model provides situational awareness by identifying regional threat hotspots. In the proposed system, interactive world maps display attack origins and targets, supporting rapid identification of high-risk geographical areas.

### **How it works:**

Maps data using latitude and longitude to analyze spatial patterns.

### **Uses in Cybersecurity:**

- Attack origin visualization
- Target region analysis

### **Accuracy:**

High when location data is available

### **Relevance to Project:**

Directly supports **geospatial risk mapping** using interactive world maps.

## **10. Visual Analytics Model**

The Visual Analytics model combines data processing with interactive visual representation to support human decision-making.

Rather than relying on automated predictions, this model allows analysts to explore data dynamically using filters, drill-downs, and comparative views. It serves as the core analytical approach of the project, enabling effective interpretation of cybersecurity data through interactive dashboards.

### **How it works:**

Combines human interaction with visual representation of data to support decision-making.

### **Uses in Cybersecurity:**

- Threat exploration
- Executive-level reporting
- Interactive filtering and drill-down analysis

### **Accuracy:**

High for insight discovery and situational awareness

### **Relevance to Project:**

Core model underpinning the entire dashboard architecture.

## **Model Selection Justification**

The primary objective of the **Interactive Cyber Threat Visualization Dashboard** is not automated threat prediction, but **clear, interpretable, and actionable visualization of cybersecurity data**. Therefore, the model selection prioritizes **explainability, real-time interaction, and analytical clarity** over complex black-box machine learning models.

Based on the project requirements and expected outcomes, the following models are most suitable:

#### **Selected Core Models**

- **Descriptive Statistical Analysis**
- **Time-Series Analysis**
- **Hierarchical Analysis**
- **Geospatial Analysis**
- **Visual Analytics Model**

#### **Justification**

- These models align directly with the project outcomes such as **trend detection, vulnerability prioritization, and geographical risk mapping**.
- They enable **real-time interactive visualization**, which is essential for analyst-driven exploration.
- The models are **data-efficient**, making them suitable for simulated datasets without requiring extensive labeled data.
- They provide **high interpretability**, which is critical for executive reporting and academic evaluation.
- Integration with **Python, Pandas, Plotly, and Dash** is seamless and computationally efficient.

#### **Why Advanced ML Models Were Not Chosen**

While models such as Random Forests, SVMs, and Neural Networks offer high predictive accuracy, they are more suitable for **intrusion detection and automated threat prediction systems**. Since this project focuses on **visual analytics rather than prediction**, such models were intentionally excluded to maintain system simplicity, transparency, and performance.

#### **Final Model Decision**

The project adopts a **Hybrid Visual Analytics–Driven Analytical Model**, combining statistical, temporal, hierarchical, and geospatial analysis to deliver an interactive and insight-rich cybersecurity dashboard.