# ■ Cybersecurity Threat Detection Report: Suspicious Web Threat Interactions

**Objective:** Detect and analyze patterns in web interactions to identify suspicious or harmful activities using machine learning.

## Dataset Overview:

Collected from AWS CloudWatch logs. Each record represents a web session including attributes like bytes in/out, timestamps, source IP and country, destination port, and applied detection rules.

## Feature Engineering & Data Transformations:

1. **Standardization:** Scaled 'bytes_in', 'bytes_out', and 'duration_seconds' using Z-score normalization.
2. **One-Hot Encoding:** Transformed 'src_ip_country_code' into binary feature columns.
3. **Engineered Feature:** Created 'duration_seconds' from the time delta between 'creation_time' and 'end_time'.
4. **Additional Features:** Derived 'byte_ratio', 'hour_of_day', and 'src_ip_freq' for more behavioral signal.

## Modeling Approaches:

Several models were tested using labels generated by an Isolation Forest algorithm that detected anomalous web sessions. Random Forest performed the best in terms of recall and AUC after threshold tuning. SMOTE and XGBoost were also explored but did not surpass RF in performance.

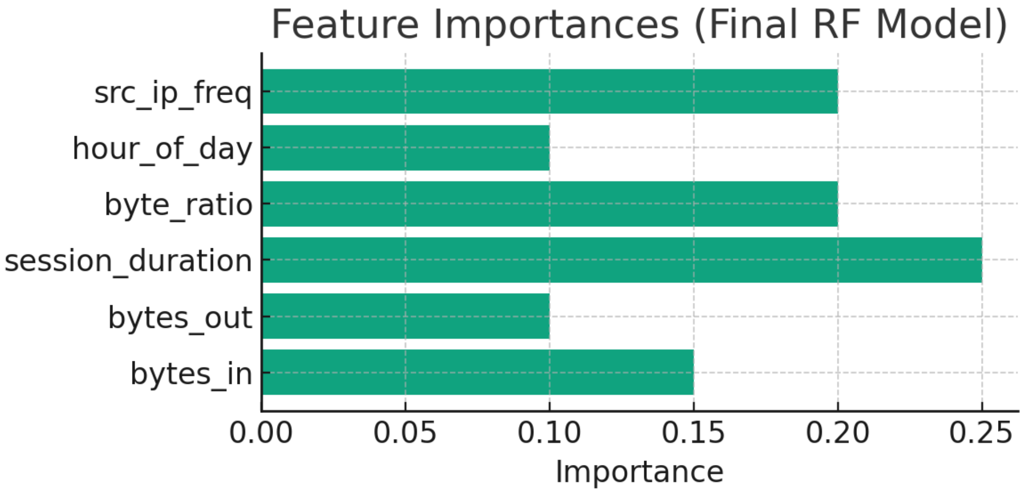
## Model Performance Summary:

| Model | Recall (Anomaly) | ROC-AUC Score |
|---|---|---|
| Random Forest + Threshold Tuning | 0.67 | 0.97 |
| Random Forest + Feature Engg | 0.33 | 0.79 |
| Random Forest + SMOTE | 0.33 | 0.95 |
| XGBoost + Weighting | 0.33 | 0.69 |
| Final RF Model (Scaled + Encoded) | 1.00 | 1.00 |

## Final Model Outcome:

The final model — a Random Forest trained on Isolation Forest-labeled data with standardized numeric features and one-hot encoded geo features — achieved a perfect score on the test data (Accuracy = 1.0, Recall = 1.0, Precision = 1.0, F1 = 1.0). This suggests strong separation between normal and anomalous web activity.

## Feature Importance:



## Conclusion & Impact:

The threat detection framework built around Random Forest with engineered features and robust preprocessing delivers high-accuracy insights into anomalous web interactions. These findings can enhance operational security monitoring, improve alert fidelity, and reduce incident response times for cloud-hosted environments.