09-11-2025

ANOMALY DETECTION AND ATTACK PREDICTION FROM STATEFUL FIREWALL LOGS

TEAM - 8

- AKILAN RAJA (RA2512049015004)
- THARUN DANANJAI (RA2512049015003)
- IMRAN KHAN (RA2512049015052)
- ASHWIN BALAJI RAMESH (RA2512049015006)

INTRODUCTION

- **Network security is a growing challenge** as modern IT infrastructures generate massive volumes of firewall logs that capture every network connection, packet, and policy decision.
- **Manual analysis of these logs is impractical** — traditional rule-based systems struggle to detect unknown or evolving cyber-attacks hidden within complex traffic patterns.
- This project aims to **automate anomaly detection and attack prediction** by applying **machine learning techniques** to analyze stateful firewall logs and identify suspicious or abnormal network behavior.
- Through **Exploratory Data Analysis (EDA)**, the project uncovers key patterns, detects outliers, and visualizes traffic characteristics to understand the relationship between network features and potential attacks.
- By combining **supervised and unsupervised learning models**, the system can both **detect new anomalies** and **predict known attack types**, enabling faster, smarter, and more reliable network threat detection.

PROBLEM STATEMENT

- **Massive firewall log data** makes manual monitoring and threat detection nearly impossible in modern networks.
- **Traditional rule-based systems** fail to detect new or evolving cyber-attacks due to reliance on static signatures.
- There is a **lack of intelligent automation** to analyze firewall logs and identify anomalies in real time.
- The challenge is to **develop a machine learning–based system** that can detect abnormal traffic patterns and accurately predict potential attacks.

OBJECTIVE

Objective 1:
To **analyze and preprocess firewall logs** by removing duplicates, handling missing values, and identifying outliers to ensure data quality for machine learning analysis.

Objective 2:
To perform **Exploratory Data Analysis (EDA)** for understanding traffic behavior, visualizing feature relationships, and detecting patterns indicative of normal or malicious activity.

Objective 3:
To develop and compare **supervised and unsupervised machine learning models** for effective anomaly detection and attack prediction.

Objective 4:
To train the chosen Supervised (Classification) and Unsupervised(Anomaly) models based on the performance metrics.

Objective 5:
To **evaluate model performance** using metrics such as accuracy, precision, recall, F1-score, and AUC, ensuring reliable detection of both known and unknown attacks for proactive network defense.

LITERATURE REVIEW

- Machine learning methods are increasingly used in intrusion detection systems to move beyond static signature-based rules. [https://www.mdpi.com/1424-8220/23/5/2415]
- The Isolation Forest algorithm is widely applied for anomaly detection in network traffic due to its efficiency and unsupervised nature. [https://www.mdpi.com/2227-9709/11/4/83]
- Studies show firewall log-based anomaly detection benefits from combining feature engineering (timestamps, byte/packet ratios) with ML algorithms. [https://www.ncbi.nlm.nih.gov/articles/PMC11054483]
- Deep learning and sequence modelling are emerging in IDS research to capture temporal and contextual dependencies in log data. [https://arxiv.org/abs/2504.07839]
- A key challenge is class imbalance and the scarcity of labeled attack data, which limits supervised model performance in real-world networks. [https://arxiv.org/abs/2009.07352]
- High-cardinality categorical features (e.g., source IPs, services) in firewall logs require thoughtful encoding strategies to avoid model overfitting and memory explosion. [https://www.mdpi.com/2073-8994/14/12/2668]
- Real-time anomaly detection for streaming log data is gaining importance, with unsupervised models offering scalable solutions for large-scale network monitoring. [https://www.digitalocean.com/community/tutorials/anomaly-detection-isolation-forest]

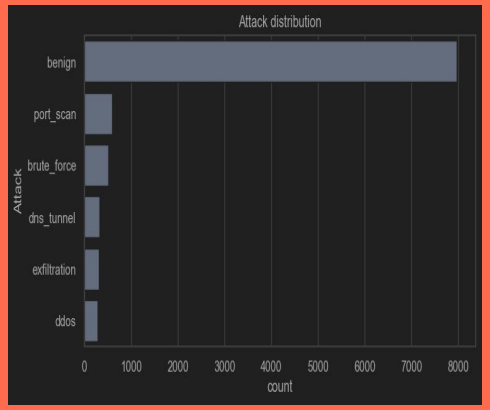
DATA ACQUISITION AND PRE-PROCESSING

- **Firewall log data** was collected from UCI Machine Learning Repository containing key attributes such as IP addresses, ports, protocols, bytes, packets, duration, and actions.
- The dataset was **imported and inspected** for missing values, duplicates, and inconsistent data types to ensure data integrity.
- **Data cleaning** included removing duplicate rows, coercing invalid values, and handling missing entries — numeric fields were filled with the median, and categorical fields with mode or forward/backward fill.
- **Feature engineering** extracted useful patterns such as hour, weekday, and private/public IP indicators, while converting IPs to integer representations.
- **Outlier detection and transformation** (using IQR and log-scaling) were applied to stabilize skewed features like *Bytes*, *Packets*, and *Duration* for better model performance.

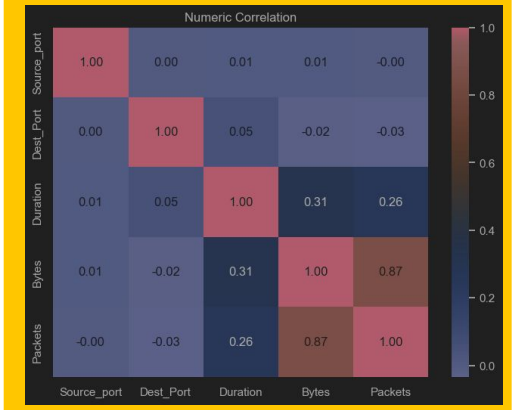
EXPLORATORY DATA ANALYSIS

- Data is **clean and consistent**, with no missing or duplicate records.
- **Outliers are meaningful** and represent potential anomalies or high-traffic attack events.
- **Strong correlation** exists between bytes and packets, reflecting logical traffic patterns.
- **Attack distribution is imbalanced**, which must be handled during model training.
- **Protocol, Service, Duration, and Bytes** are key features that strongly influence anomaly and attack detection.

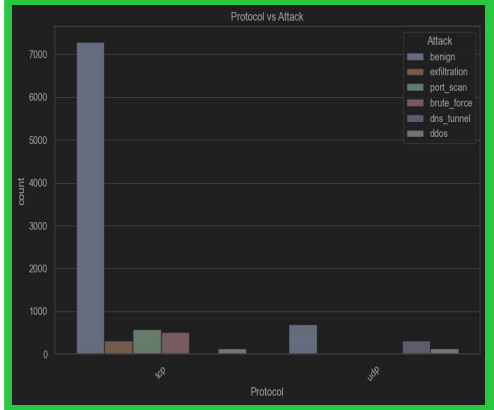
EXPLORATORY DATA ANALYSIS



Attack Distribution



Numeric Correlation



Protocol Vs Attack

MODEL SELECTION

1. **Supervised Learning**
    - In the **supervised learning** evaluation, multiple classifiers were tested, and **Gradient Boosting** outperformed others with the highest **F1-score of 0.89**, showing excellent accuracy and balanced precision–recall in predicting attack types.
    - **Chosen Algorithm:** *Gradient Boosting*.
2. **UnSupervised Learning**
    - In the **unsupervised learning** evaluation, **Isolation Forest** achieved the best **AUROC of 0.75**, effectively distinguishing anomalous network behaviors without labeled data.
    - **Chosen Algorithm:** *Isolation Forest.*

SUPERVISED LEARNING SCORES

Winner : Gradient Boosting

Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.9211	0.8482	0.7381	0.7699
Decision Tree	0.945	0.8778	0.8778	0.8643
Random Forest	0.95	0.9284	0.8183	0.8587
Gradient Boosting	**0.9571**	**0.9229**	**0.8692**	**0.8906**
SVM	0.9011	0.7220	0.6501	0.6660

UNSUPERVISED LEARNING SCORES

Winner : Isolation Forest

Model	AUROC	AP	F1
Isolation Forest	**0.7483**	**0.4108**	**0.4886**
One-Class SVM	0.7177	0.4001	0.4345
Local Outlier Factor	0.5473	0.2723	0.3433

FLOW DIAGRAM

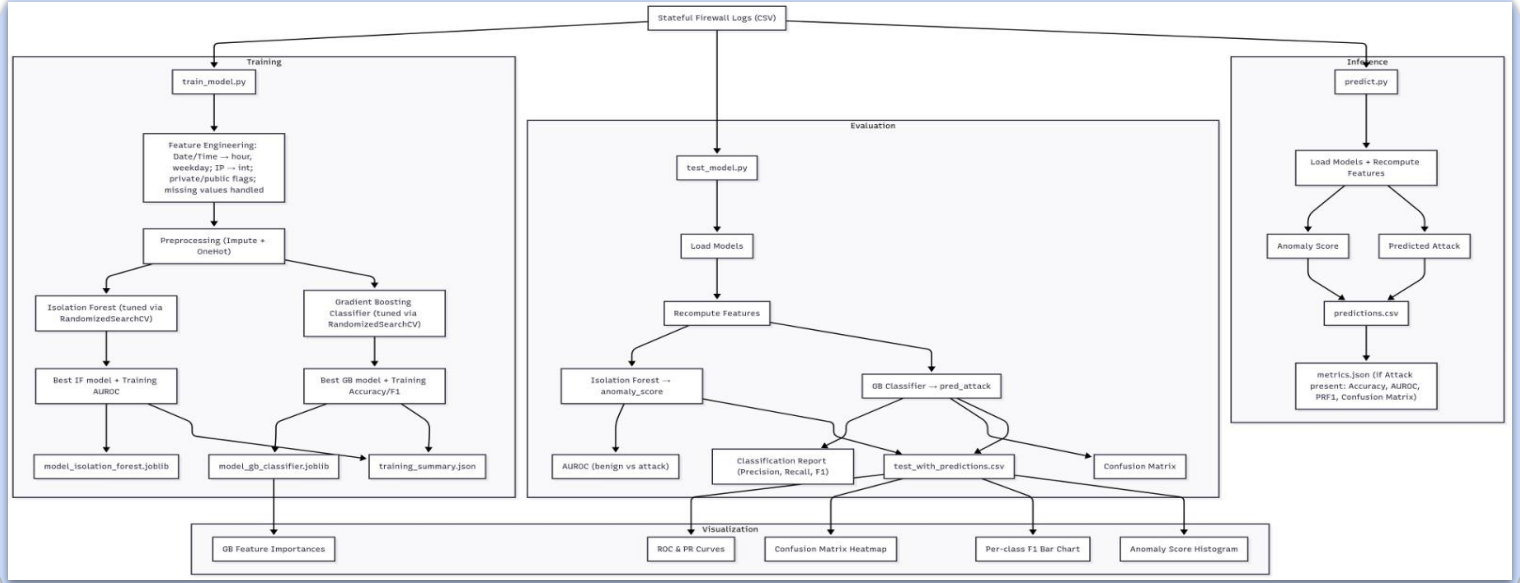


Fig. 1 Train and Evaluate the Model

FLOW DIAGRAM

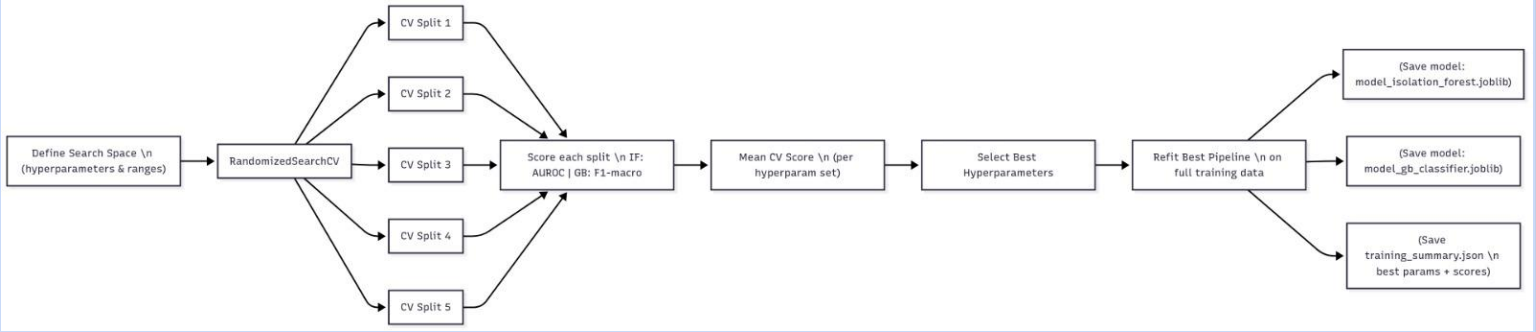


Fig. 2 Randomised CV Search

TRAINING THE MODEL

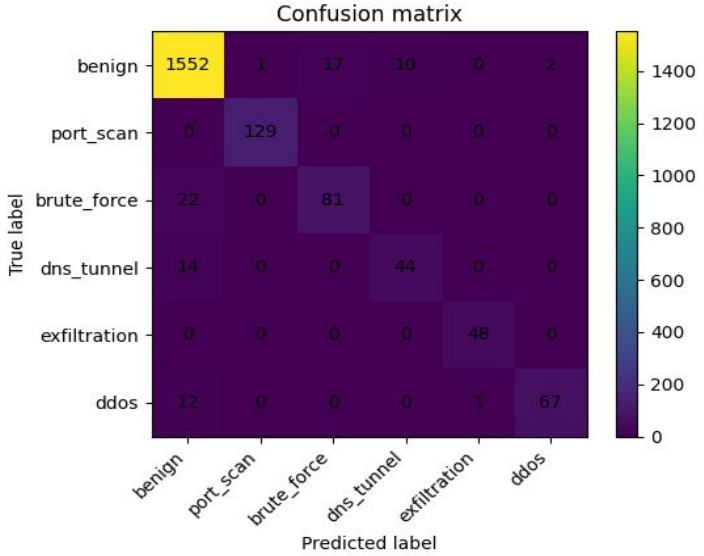
- The dataset was preprocessed, cleaned, and transformed using feature engineering before training both supervised and unsupervised models.
- **Gradient Boosting Classifier** (supervised) achieved the highest **F1-score of 0.90**, showing strong predictive accuracy in attack classification.
- **Isolation Forest** (unsupervised) reached an **AUROC of 0.79**, effectively detecting anomalies without labeled data.
- Both models were fine-tuned using **cross-validation** and **RandomizedSearchCV** for optimal performance.
- Final models were saved for deployment — **Gradient Boosting** for attack prediction and **Isolation Forest** for anomaly detection.

EVALUATING THE MODEL

- The trained models were evaluated using a separate test dataset to measure real-world accuracy and reliability.
- **Gradient Boosting Classifier** achieved **96% accuracy** and a **macro F1-score of 0.91**, proving strong attack-prediction capability.
- **Isolation Forest** achieved an **AUROC of 0.79** and **Average Precision of 0.48**, effectively identifying anomalies in unlabeled data.
- Both models demonstrated balanced precision and recall, though minor confusion occurred among similar attack types.
- Overall, the hybrid approach ensures robust **anomaly detection** and **attack classification** across varied network traffic.

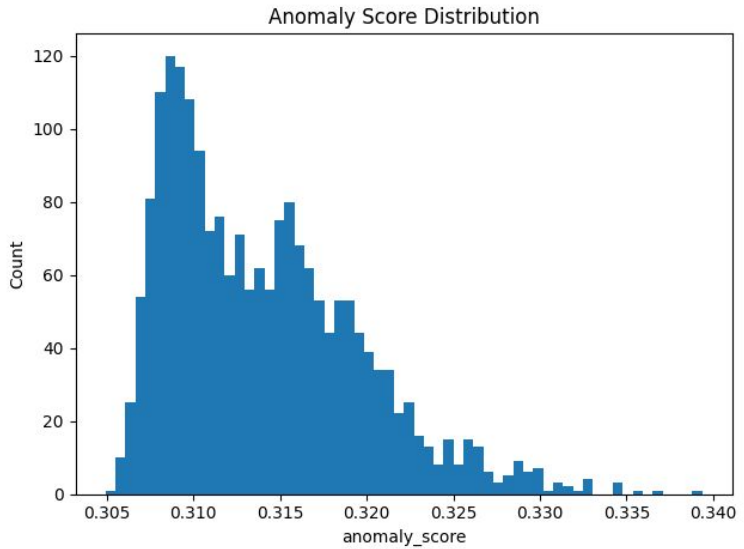
Confusion Matrix

- Shows actual vs predicted attack classes by the Gradient Boosting model.
- Most diagonal values are high, proving accurate classification for major attack types.
- Minor misclassifications occur between **`brute_force`**, **`dns_tunnel`**, and **`benign`.**
- Perfect detection for **`port_scan`** and **`exfiltration`** demonstrates strong model focus.
- Visual proof that the model handles multiple attack categories effectively.



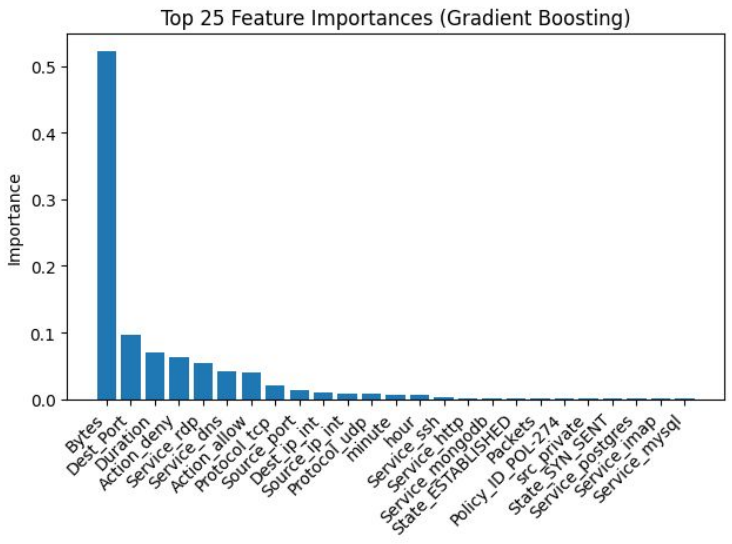
Anomaly Score

- Displays the frequency of anomaly scores predicted by the **Isolation Forest model**.
- Most values cluster around 0.31, indicating normal traffic concentration.
- The right-hand tail represents potential anomalous or abnormal connections.
- A smooth declining curve shows few extreme outliers — typical in real network data.
- Helps define threshold values to flag top-scoring anomalies for security alerts.



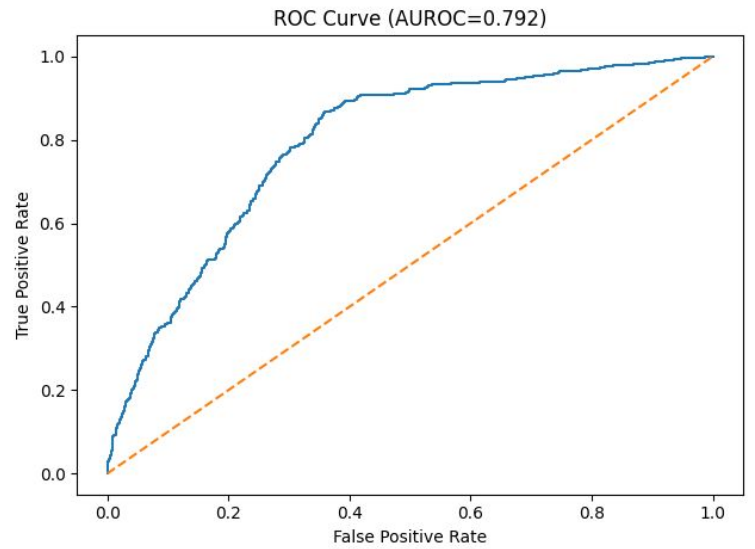
Feature Importance

- Highlights which features most influenced the Gradient Boosting classifier.
- **Bytes**, **Destination Port**, and **Duration** dominate as key predictors.
- `Action` and `Service` attributes also significantly impact classification.
- Lesser importance of `Protocol` and `Policy_ID` suggests redundancy.
- Confirms that traffic volume and session characteristics drive attack detection.



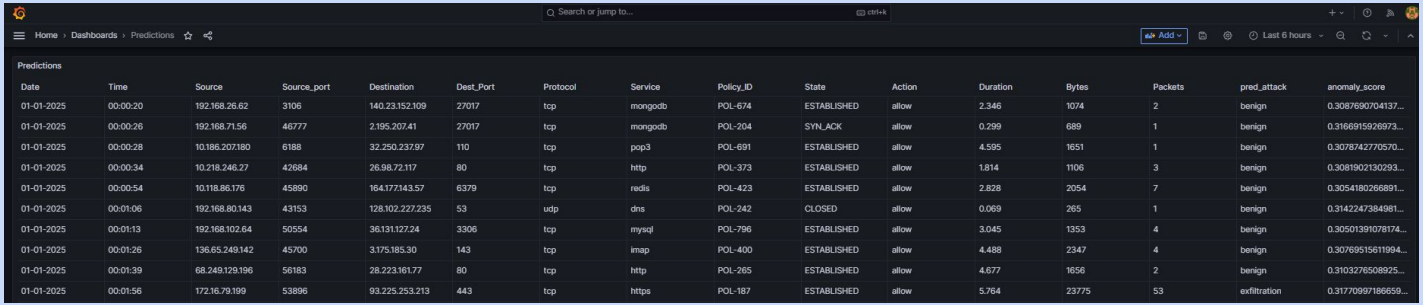
ROC Curve

- Plots **True Positive Rate vs False Positive Rate** for the Isolation Forest model.
- The **AUROC = 0.792**, meaning strong separation between normal and attack traffic.
- The curve lying above the diagonal shows good classification capability.
- Higher area under the curve indicates better anomaly ranking performance.
- Confirms the model's reliability for unsupervised anomaly detection tasks.



PREDICTIONS

- The dashboard table lists detailed network predictions from firewall log.
- Each row represents a network event with its source, destination, ports, and protocol details.
- Columns show connection attributes like Duration, Bytes, Packets, and security Action.
- pred_attack indicates the predicted classification (e.g., benign or exfiltration).
- anomaly_score quantifies how unusual each event is, with higher values signaling potential threats.

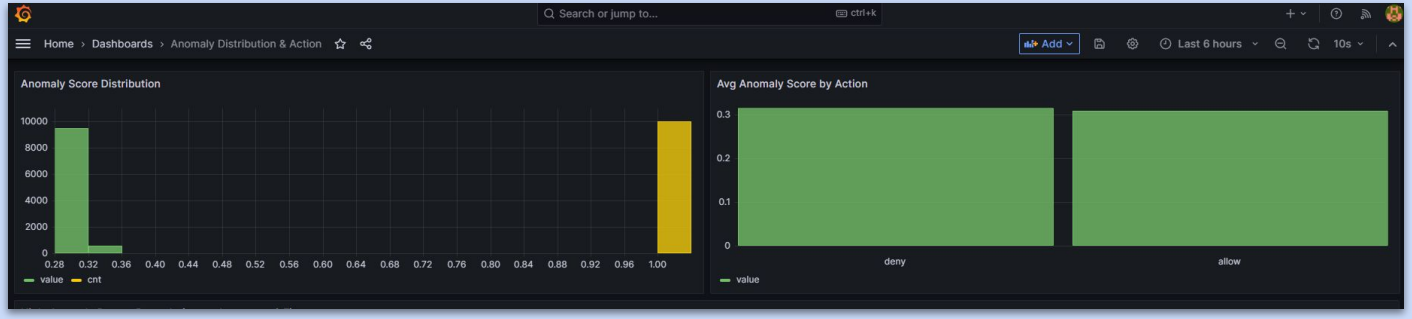


Predictions

Date	Time	Source	Source_port	Destination	Dest_Port	Protocol	Service	Policy_ID	State	Action	Duration	Bytes	Packets	pred_attack	anomaly_score
01-01-2025	00:00:20	192.168.26.62	3106	140.23.152.109	27017	tcp	mongodb	POL-674	ESTABLISHED	allow	2.346	1074	2	benign	0.3087690704137...
01-01-2025	00:00:26	192.168.71.56	46777	2.195.207.41	27017	tcp	mongodb	POL-204	SYN_ACK	allow	0.299	689	1	benign	0.3166915926973...
01-01-2025	00:00:28	10.186.207.180	6188	32.250.237.97	110	tcp	pop3	POL-691	ESTABLISHED	allow	4.595	1651	1	benign	0.3078742770570...
01-01-2025	00:00:34	10.218.246.27	42684	26.98.72.117	80	tcp	http	POL-373	ESTABLISHED	allow	1.814	1106	3	benign	0.3081902130293...
01-01-2025	00:00:54	10.118.86.176	45890	164.177.143.57	6379	tcp	redis	POL-423	ESTABLISHED	allow	2.828	2054	7	benign	0.3054180266891...
01-01-2025	00:01:06	192.168.80.143	43153	128.102.227.235	53	udp	dns	POL-242	CLOSED	allow	0.069	265	1	benign	0.3142247384981...
01-01-2025	00:01:13	192.168.102.64	50554	36.131.127.24	3306	tcp	mysql	POL-796	ESTABLISHED	allow	3.045	1353	4	benign	0.3050139107817 4...
01-01-2025	00:01:26	136.65.249.142	45700	3.175.185.30	143	tcp	imap	POL-400	ESTABLISHED	allow	4.488	2347	4	benign	0.3076951561199 4...
01-01-2025	00:01:39	68.249.129.196	56183	28.223.161.77	80	tcp	http	POL-265	ESTABLISHED	allow	4.677	1656	2	benign	0.3103276508925...
01-01-2025	00:01:56	172.16.79.199	53896	93.225.253.213	443	tcp	https	POL-187	ESTABLISHED	allow	5.764	23775	53	exfiltration	0.3177099718665 9...

ANOMALY DISTRIBUTION

- The dashboard visualizes anomaly score distribution and average scores by action in Grafana.
- The left histogram shows most anomaly scores clustered around 0.3, indicating normal behavior.
- A few extreme anomalies near score 1.0 suggest rare high-risk events.
- The right bar chart compares “deny” and “allow” actions, both averaging around 0.3 anomaly score.
- Overall, it shows system activity is mostly normal with isolated high-anomaly spikes.



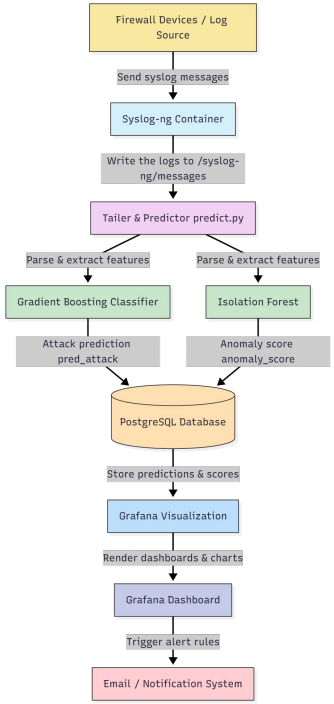
LIMITATIONS

- The dataset used is limited to a specific time frame and environment, which may not represent all real-world network behaviors.
- The **class imbalance** among attack types (e.g., few samples for `brute_force` or `dns_tunnel`) affects model generalization and detection consistency.
- **Overfitting risk** exists in the Gradient Boosting model due to perfect training scores, requiring careful validation on unseen data.
- The Isolation Forest model's performance (AUROC ≈ 0.79) shows that anomaly boundaries are not sharply defined for all traffic patterns.
- The system currently performs **batch analysis** and lacks real-time detection or continuous learning capabilities.

FUTURE ENHANCEMENTS

- Integrate **real-time log streaming and anomaly detection** using frameworks like Apache Kafka or Spark Streaming.
- Apply **deep learning approaches** (LSTM, Autoencoders, Graph Neural Networks) for sequential and contextual pattern learning from firewall events.
- Enhance data diversity by combining **multi-source logs** (firewall, IDS/IPS, system, and application logs) for holistic threat analysis.
- Implement **automated retraining pipelines** to handle evolving attack patterns (concept drift).
- Deploy the solution as a **web-based or cloud service** with real-time dashboards for monitoring, visualization, and alerting.

FUTURE STATE



- Syslog-ng collects real-time firewall logs and forwards them for processing.
- The production pipeline extracts features and applies ML models for prediction.
- Gradient Boosting predicts attack types, and Isolation Forest detects anomalies.
- PostgreSQL stores predictions and anomaly scores for visualization and analysis.
- Grafana displays real-time dashboards and triggers alerts for high-risk anomalies.

CONCLUSION

- The project developed a **machine learning-based system** to detect anomalies and predict attacks from stateful firewall logs.
- **EDA** revealed key traffic behaviors, correlations, and outliers crucial for accurate model training.
- **Gradient Boosting** achieved **96% accuracy**, while **Isolation Forest** effectively detected unseen anomalies.
- The hybrid approach improved both **attack classification** and **anomaly detection** capabilities.
- This work strengthens **network security monitoring** and sets the foundation for **real-time intelligent threat detection**.

Questions ?

THANK YOU!