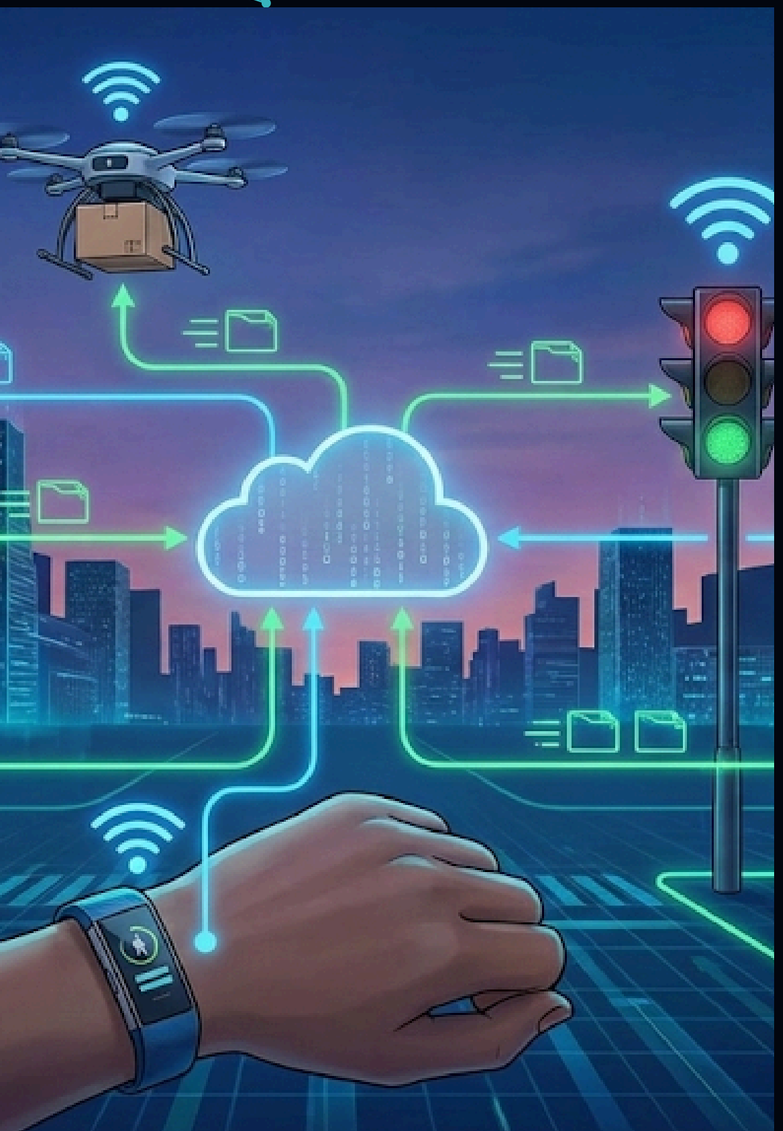
IOT BOTNET ATTACK DETECTION

using different machine learning models



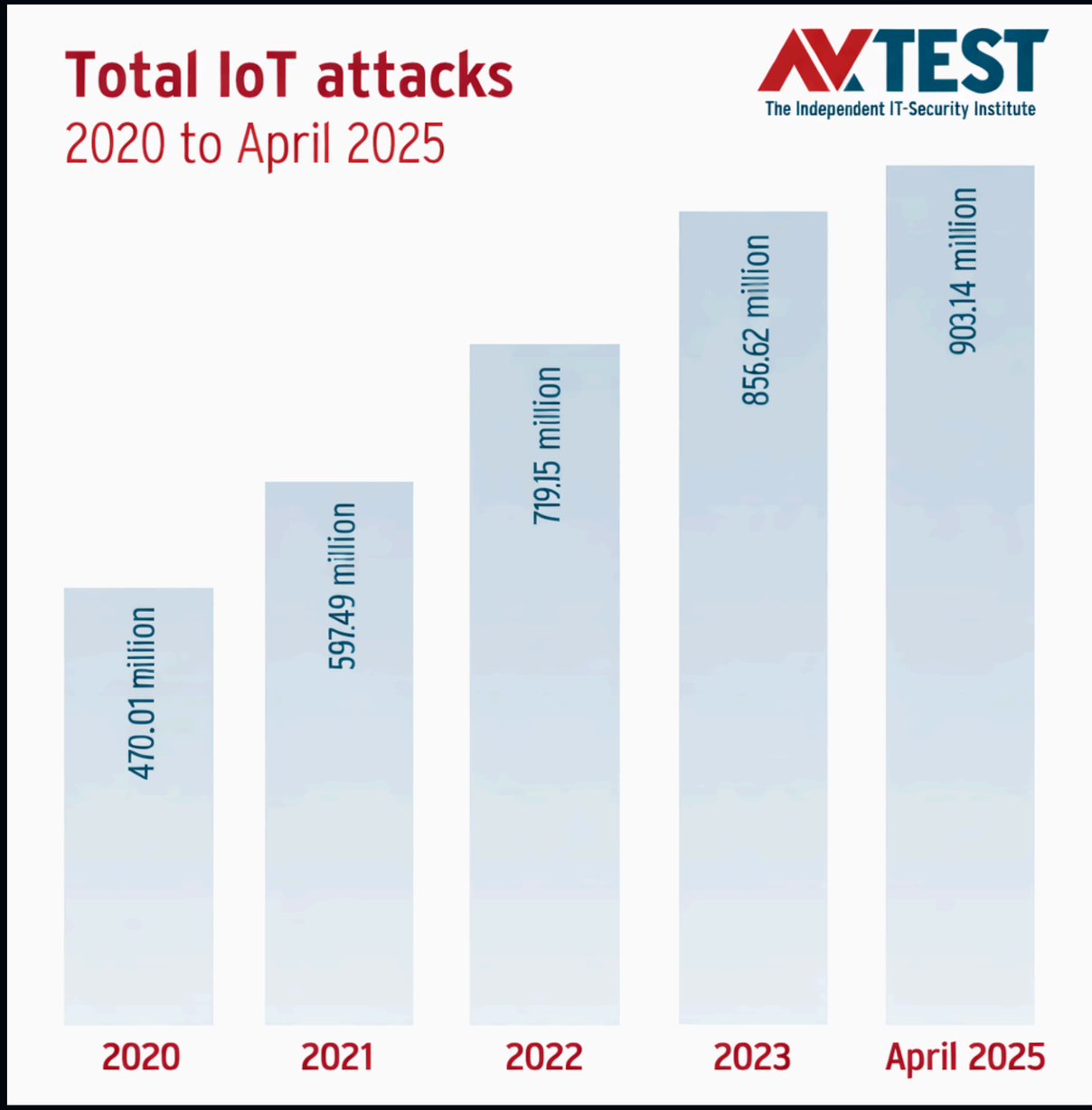
IOT

Internet of Things (IoT) is one of the most widely used cyber physical systems in the world

IoT devices are widely used in daily life

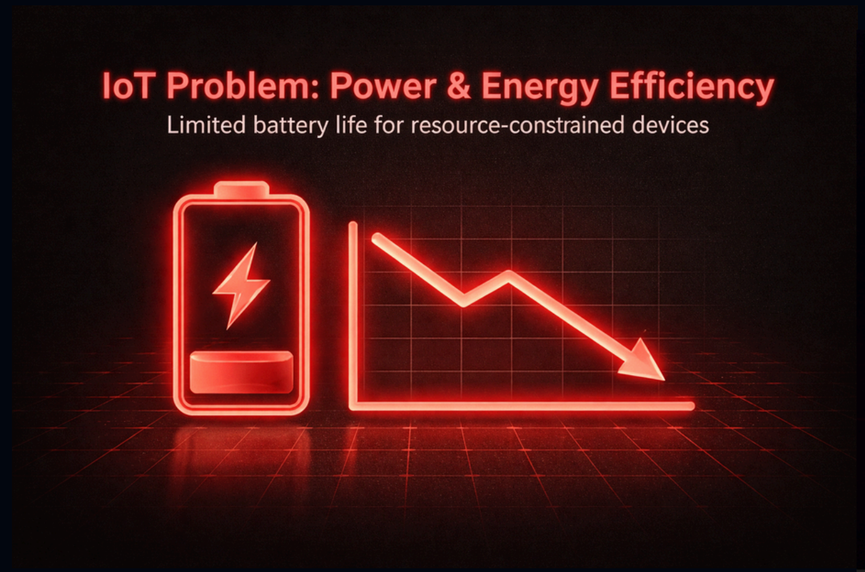
Smart homes, cameras, and sensors are everywhere

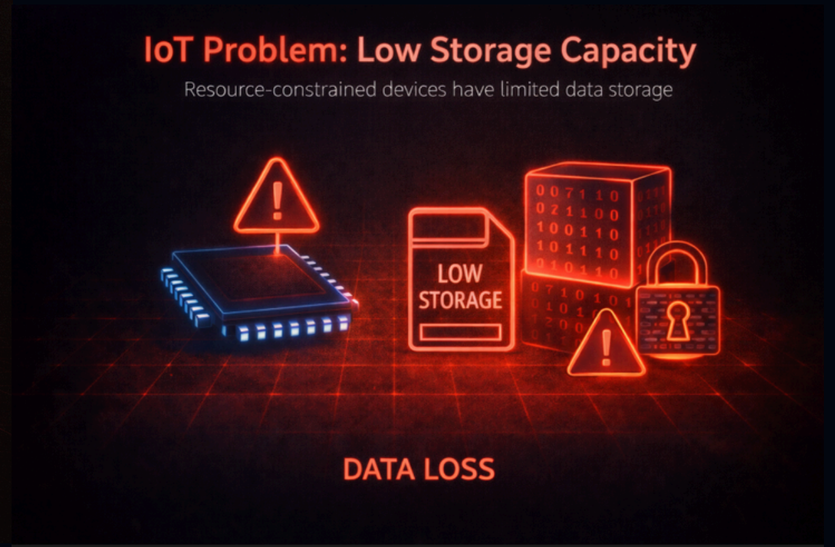
IOT are connected to the internet continuously

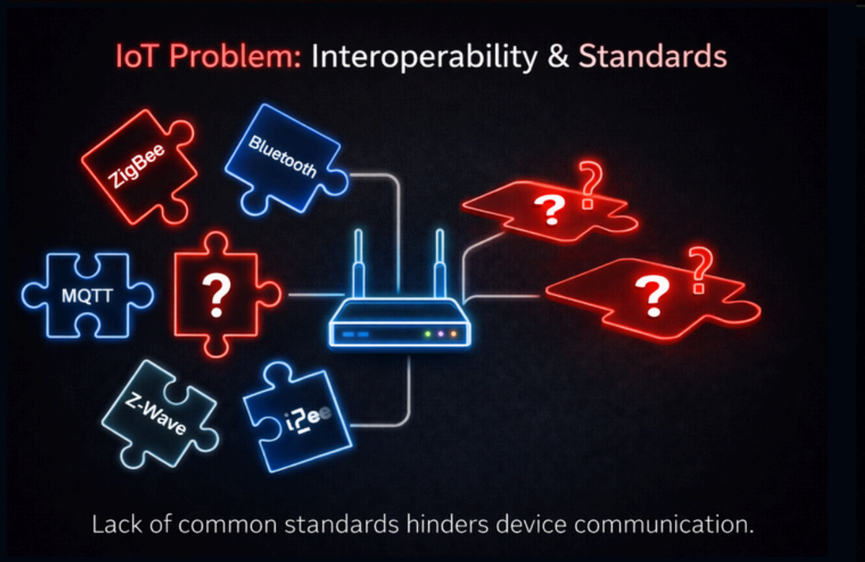


PROBLEM & MOTIVATION
WHY IOT ?

- low power
- low storage
- interoperability
- Difficult to apply traditional security





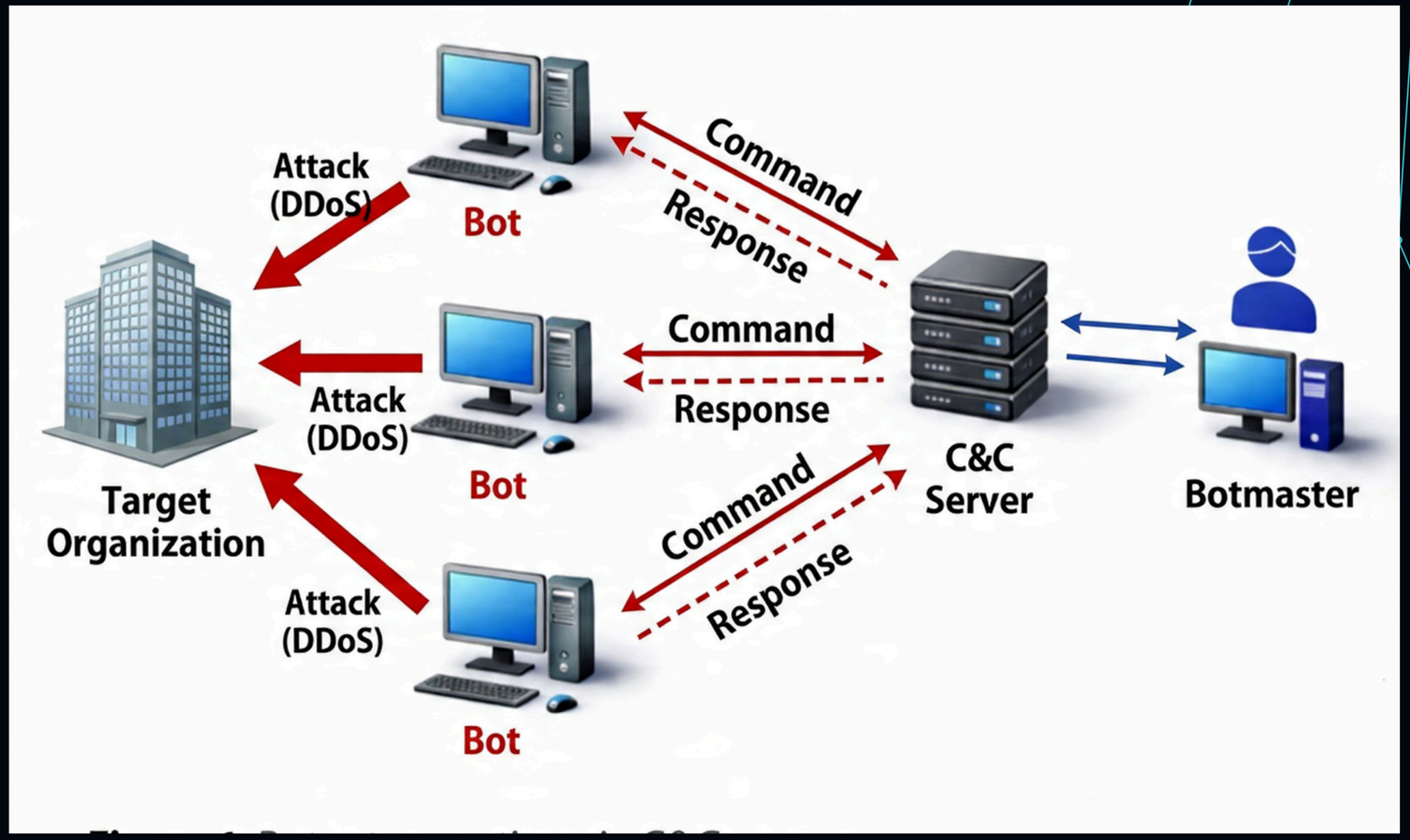


HOW A BOTNET WORKS

C&C server → sends commands

Bots → execute attacks

Target → receives DDoS traffic



DATASET DESCRIPTION

Dataset Used: N-BaIoT 2021 Dataset.

Device Types: 5 IoT devices

Traffic Types:

- **Benign: Normal device behavior.**
- **Malicious: Botnet attacks including Mirai and Bashlite families.**

WORKFLOW OVERVIEW

Proposed Workflow:

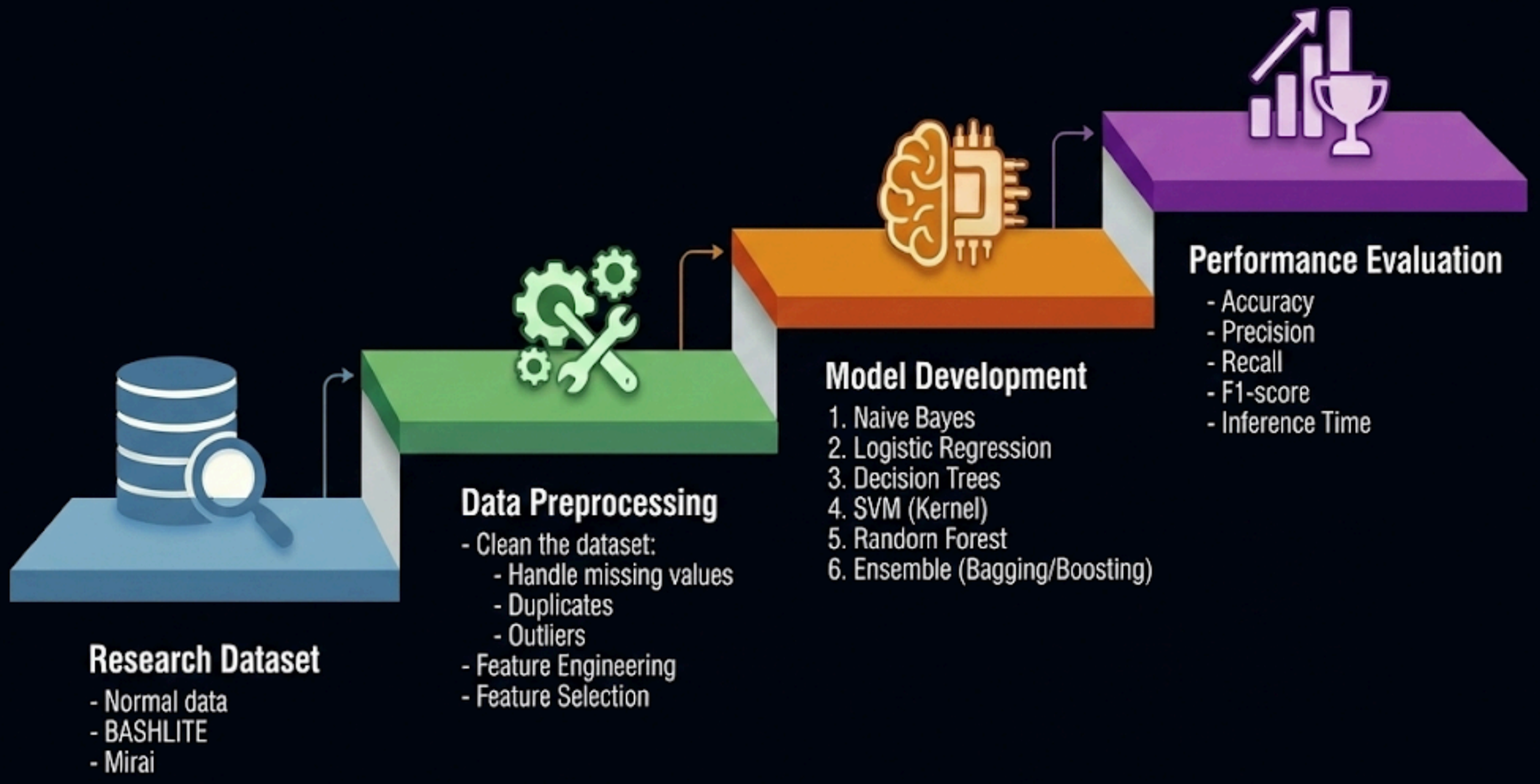
Collect IoT traffic from N-BaIoT dataset

Data preprocessing and feature selection

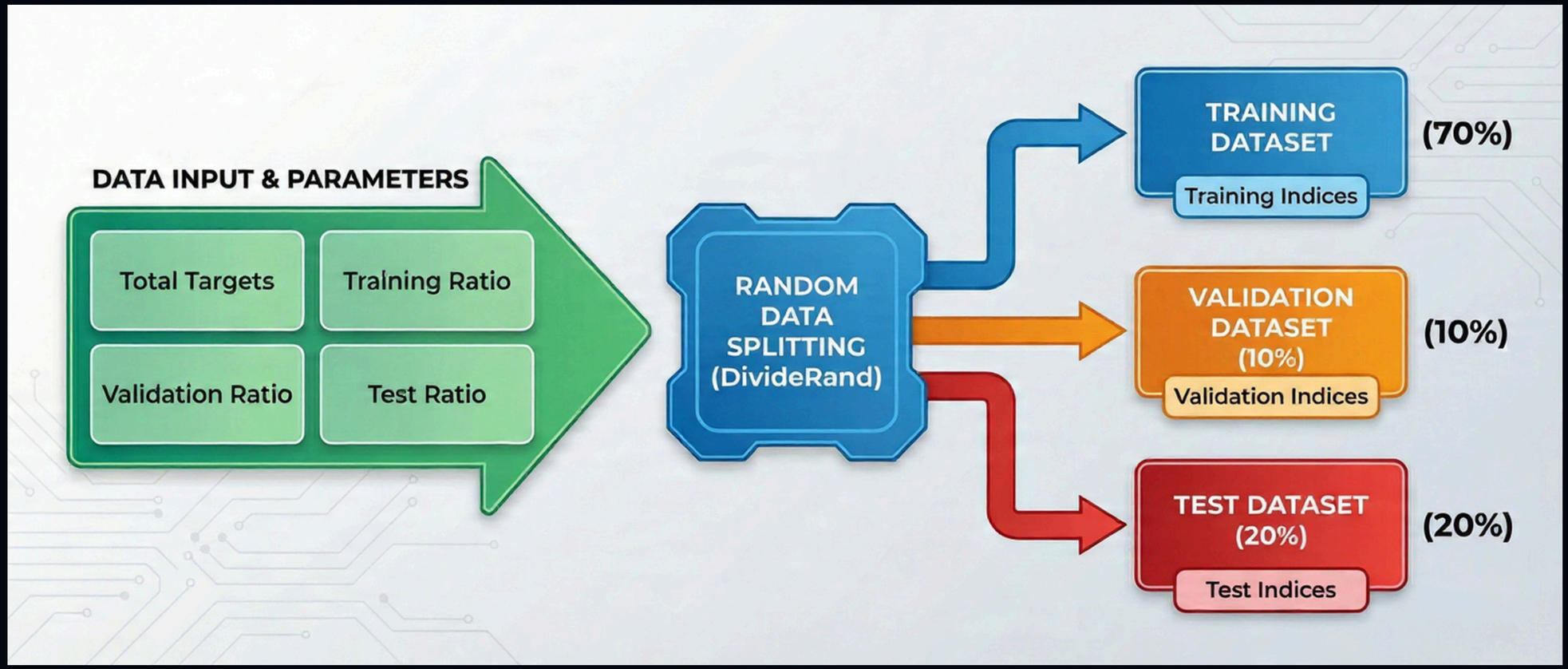
Split data into train, validation, and test sets

Train supervised machine learning models

Evaluate and compare model performance



RANDOM DATA SPLITTING



Dataset splitting strategy :

Dataset is divided into Training(70%), Validation (10%), and Testing (20%)

Training set is used for model learning

Validation set is used for tuning

Testing set is Used only for final model evaluation

DATA PREPROCESSING

Scrambled Data → Cleaning → normalization → ML-Ready Data

Data preprocessing steps:

1) Remove Duplicates and invalid records

2) Handle missing values

3) Normalize numerical features

4) Prepare Data for machine learning models

Why Normalization is Needed

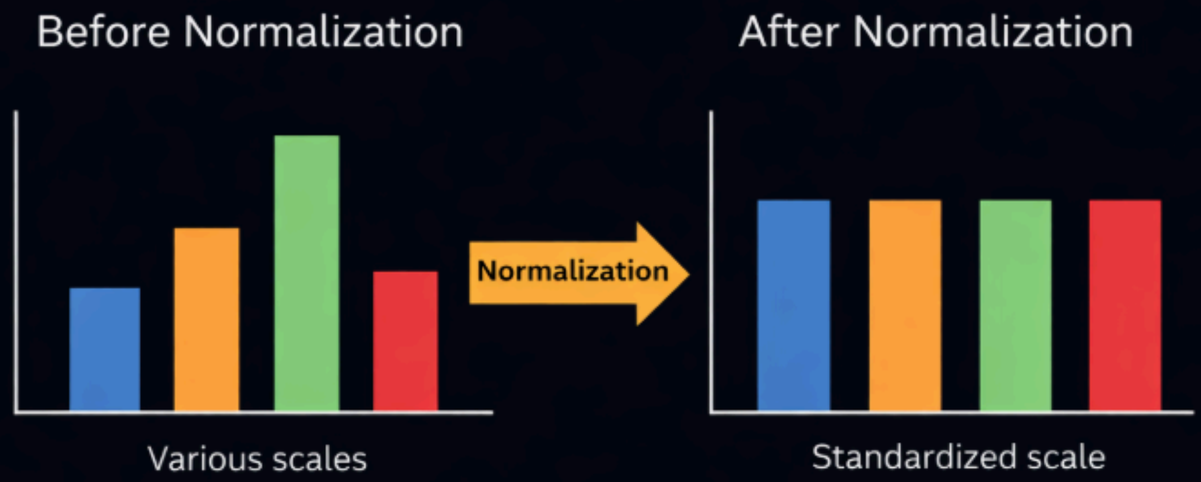
Make the process easier for the machine learning models

Prevents features with large values from affecting the final results

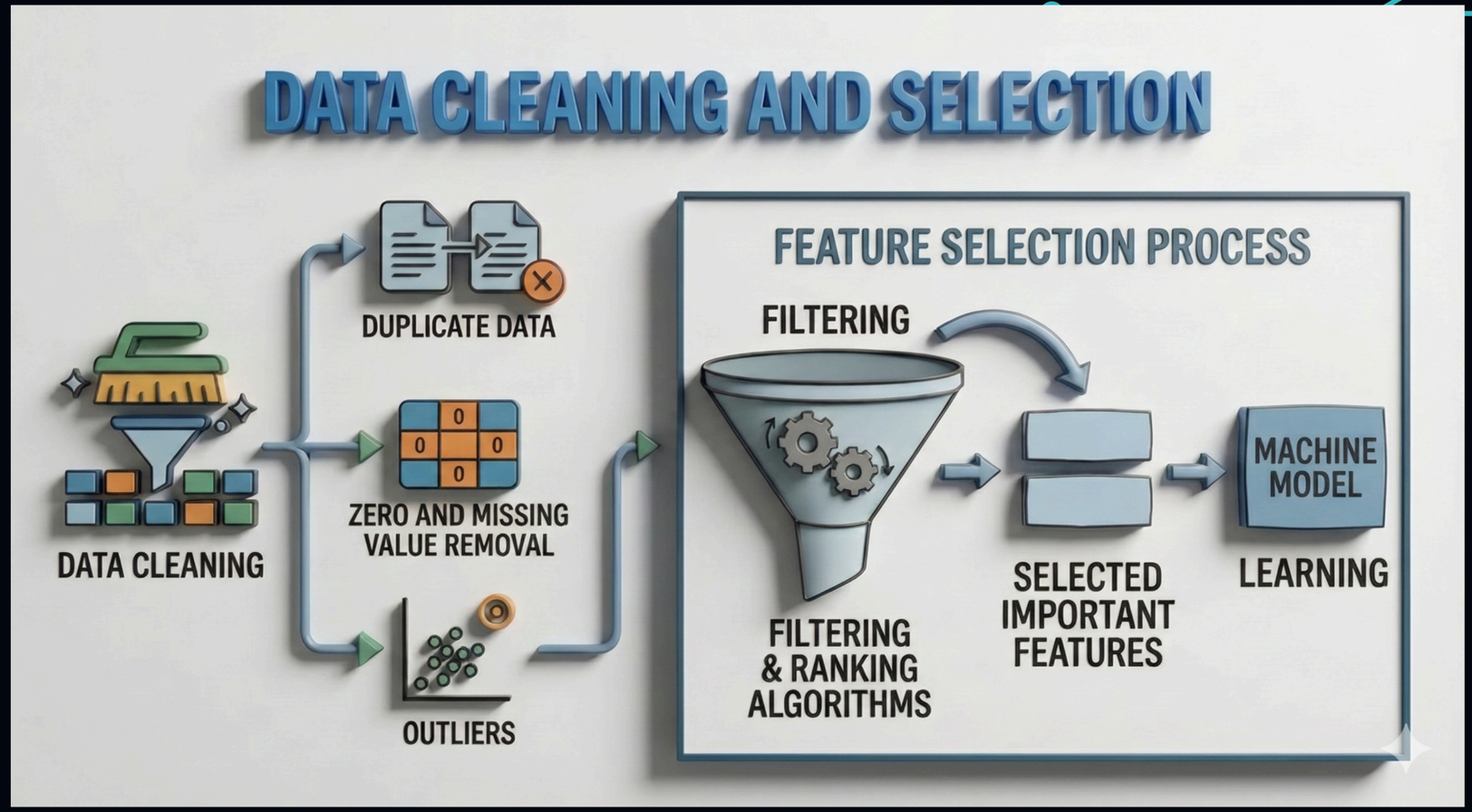
Make sure features are within the same scale

Method Used
Standardization (Z-score normalization)

$$z = \frac{x - \mu}{\sigma}$$



FEATURE SELECTION



300000 samples of normal (benign) ✓ → 284,391 normal(benign) samples.

7737 samples of malicious traffic → 7,683 botnet samples

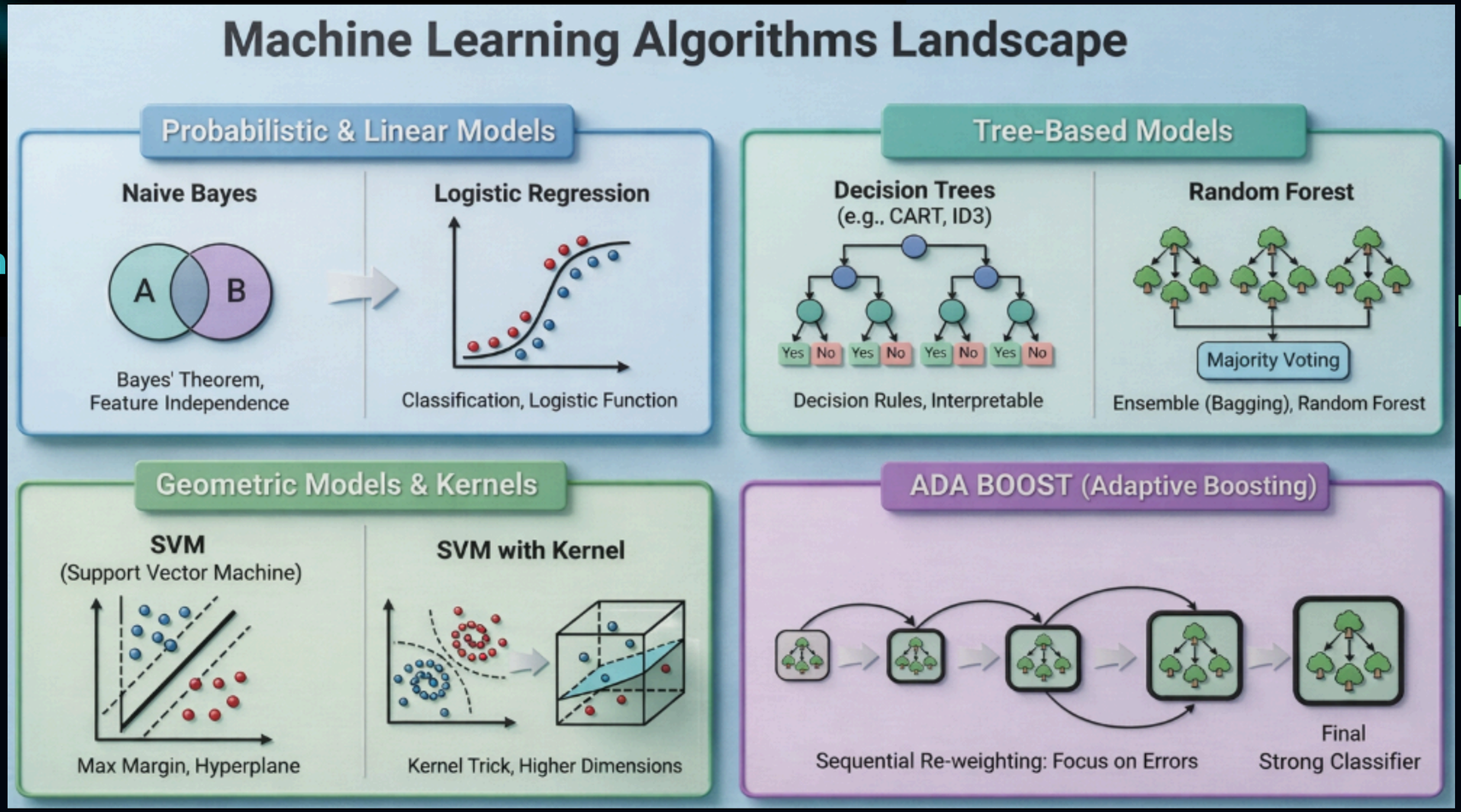
MACHINE LEARNING MODELS

The following models were implemented and compared:

Naive Bayes

Logistic Regression

SVM



Decision Tree

Random Forest

AdaBoost

Testing Phase Measurements

Accuracy

Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Precision

$$= \frac{TP}{TP + FP}$$

Recall

Recall

$$= \frac{TP}{TP + FN}$$

F1-score

F1-score

$$\text{F1-score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Inference Time

Inference Time

TP: True Positive, **TN:** True Negative, **FP:** False Positive, **FN:** False Negative

CLASSIFICATION STRATEGY

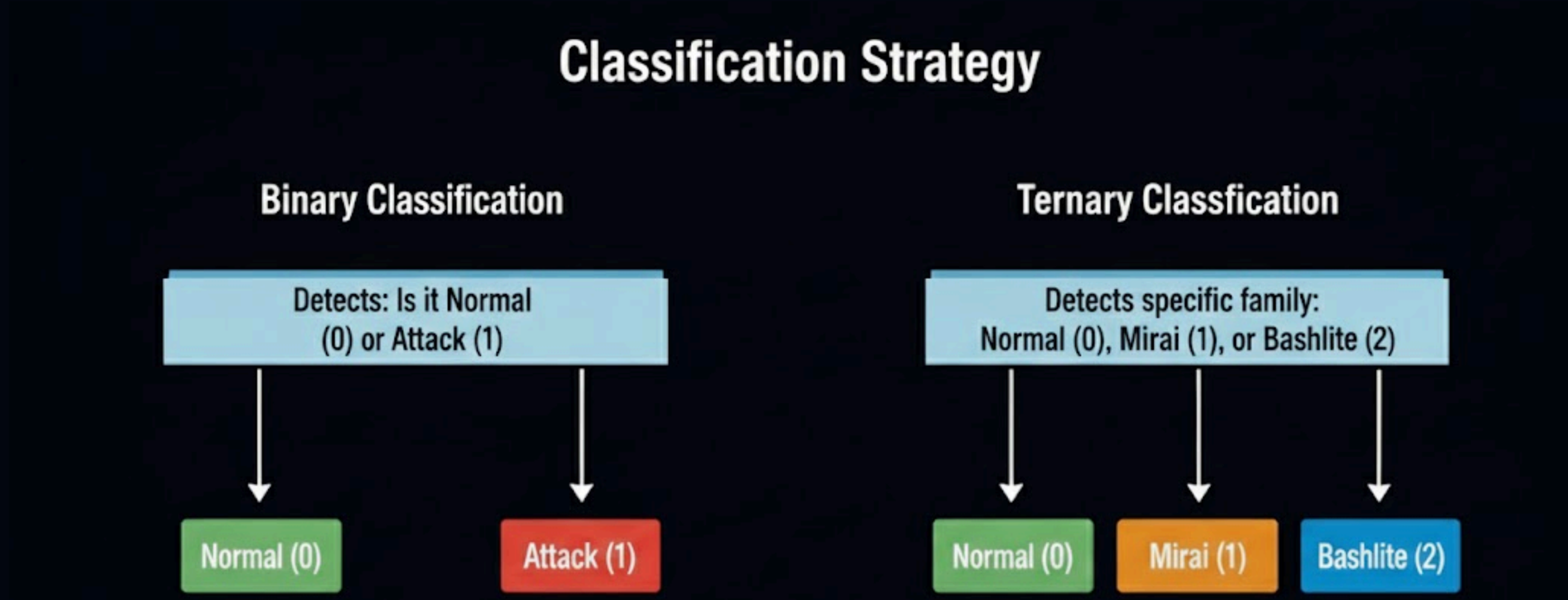
We evaluated models using two different classifiers:

Binary Classification:

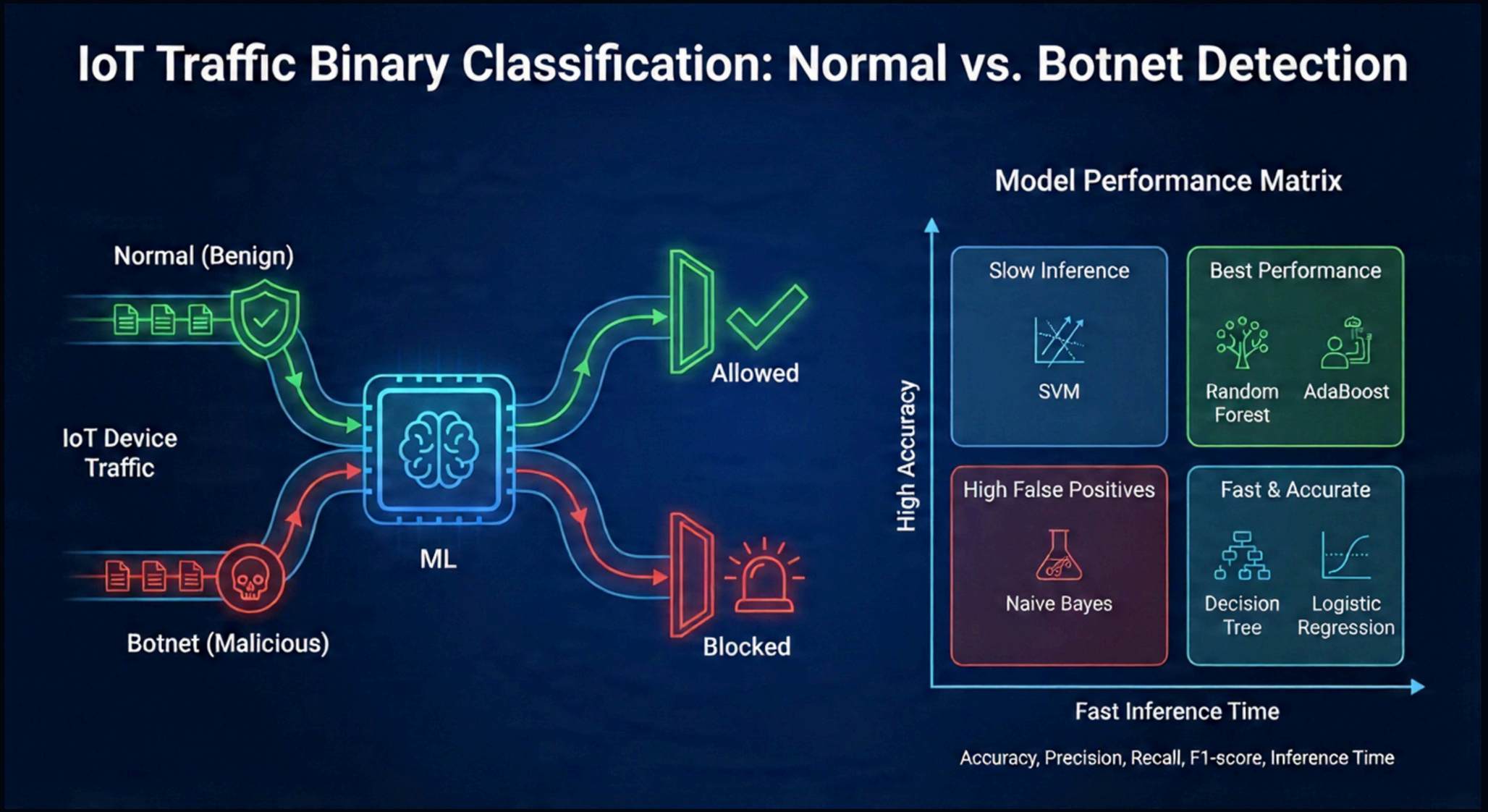
- Detects: Is it Normal (0) or Attack (1)

Ternary Classification:

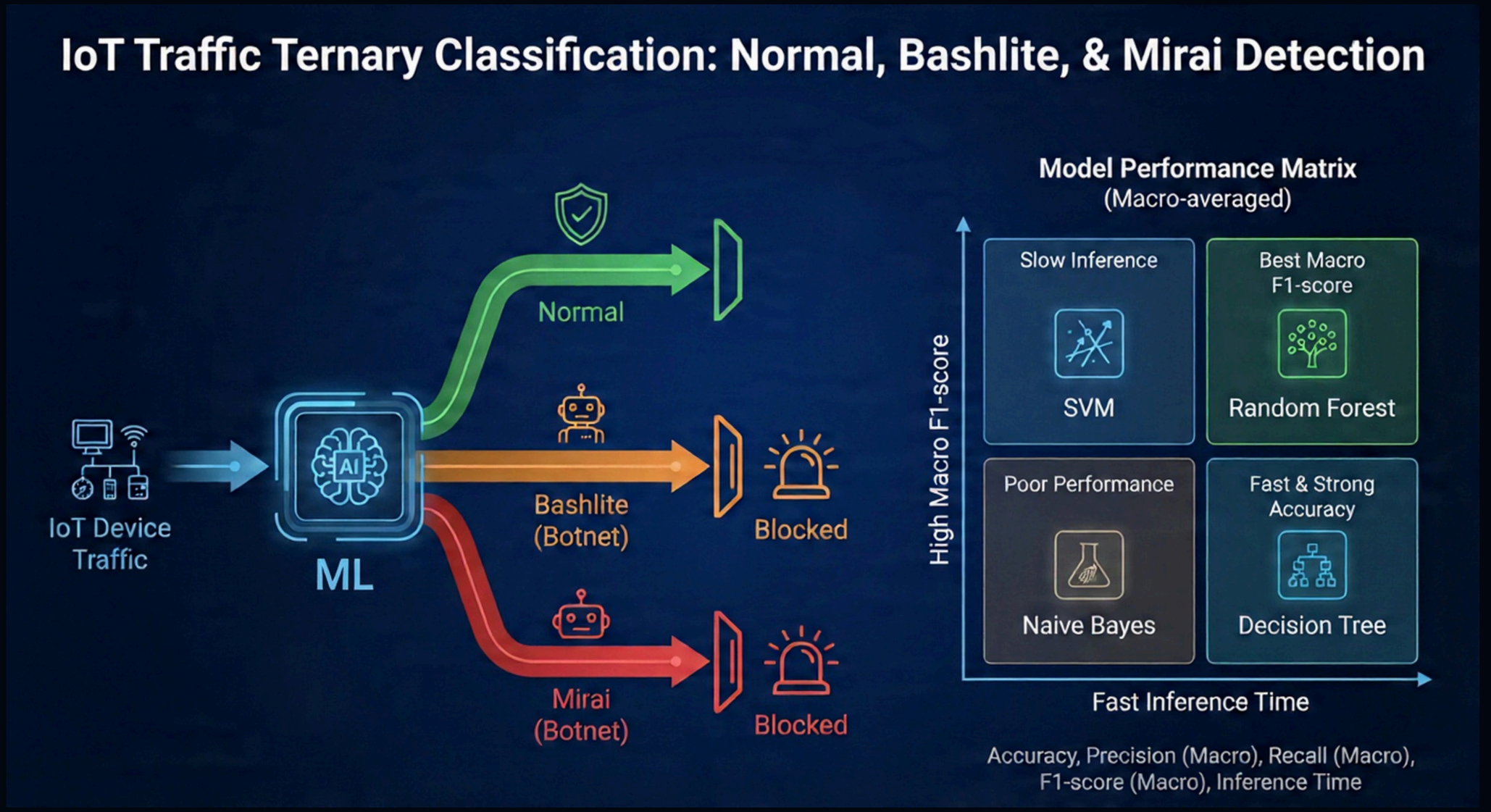
- Detects specific family: Normal (0), Mirai (1), or Bashlite (2).

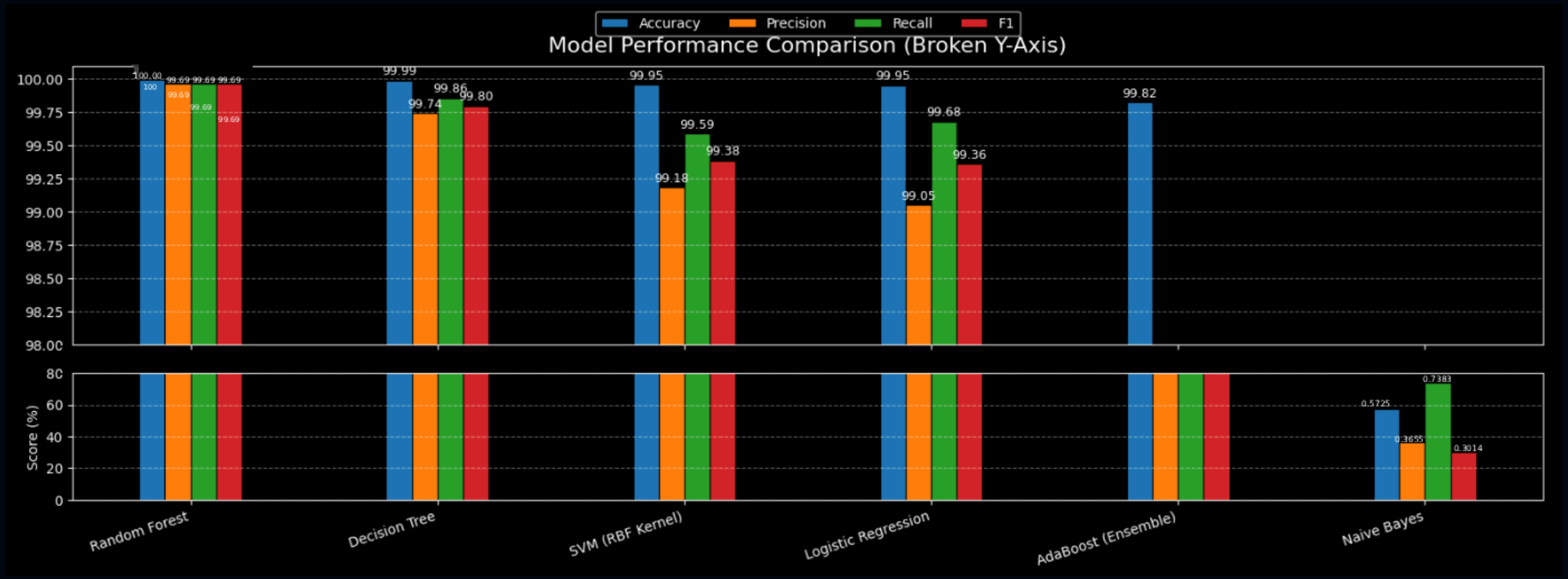


BINARY CLASSIFICATION:



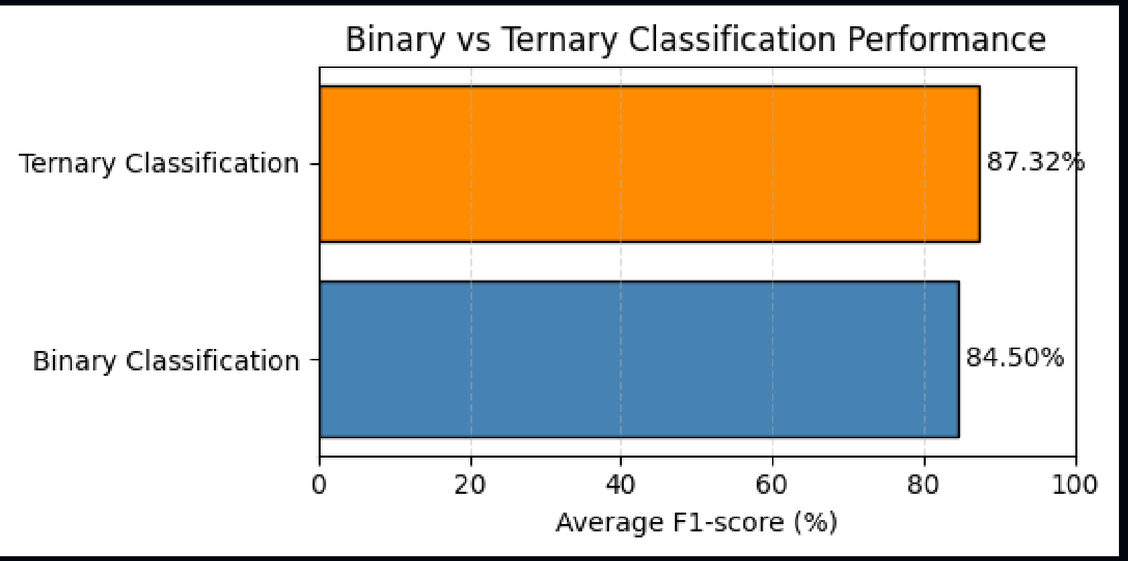
TERNARY CLASSIFICATION:



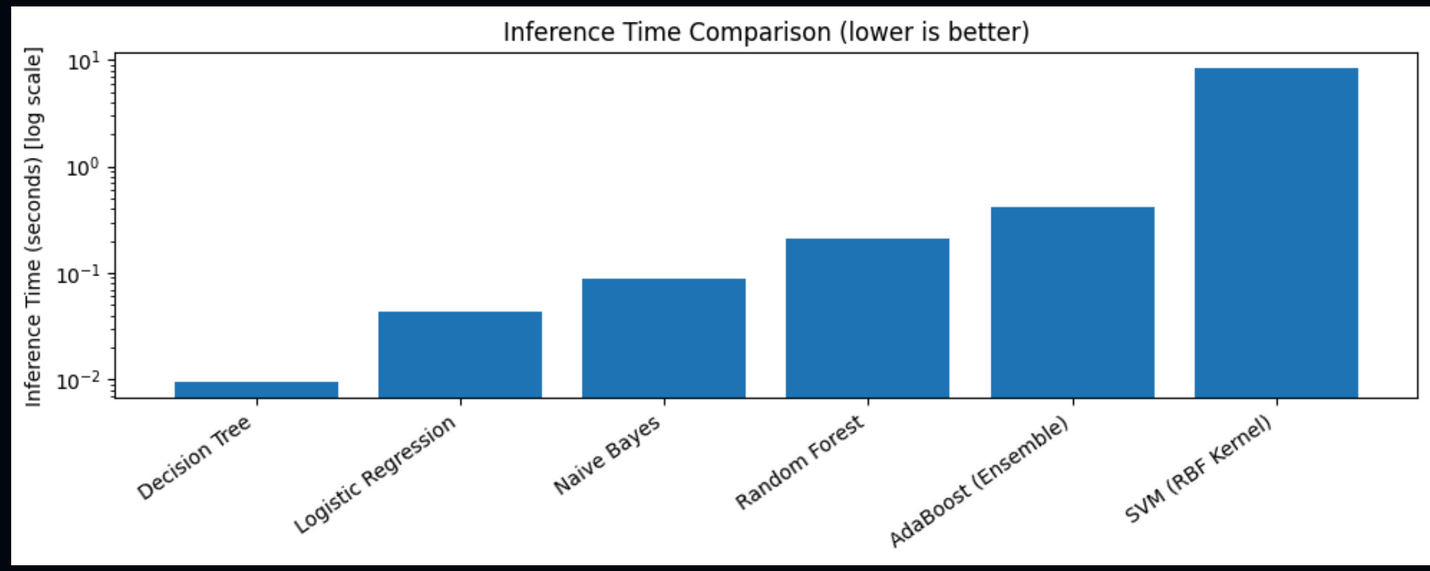

Accuracy
Precision
Recall
F1
Model Performance Comparison (Broken Y-Axis)
100.00
99.99
99.95
99.95
99.82
99.74
99.86
99.80
99.18
99.59
99.38
99.05
99.68
99.36
0.5725
0.3655
0.7383
0.3014
Score (%)
Random Forest
Decision Tree
SVM (RBF Kernel)
Logistic Regression
AdaBoost (Ensemble)
Naive Bayes

Plots

binary vs ternary performace



inference time compare



Risks of IoT (Internet of Things)

Data Privacy & Security

Unauthorized Access,
Surveillance

System Vulnerabilities

Software Bugs, Exploits,
Unpatched Devices

Physical Safety Risks

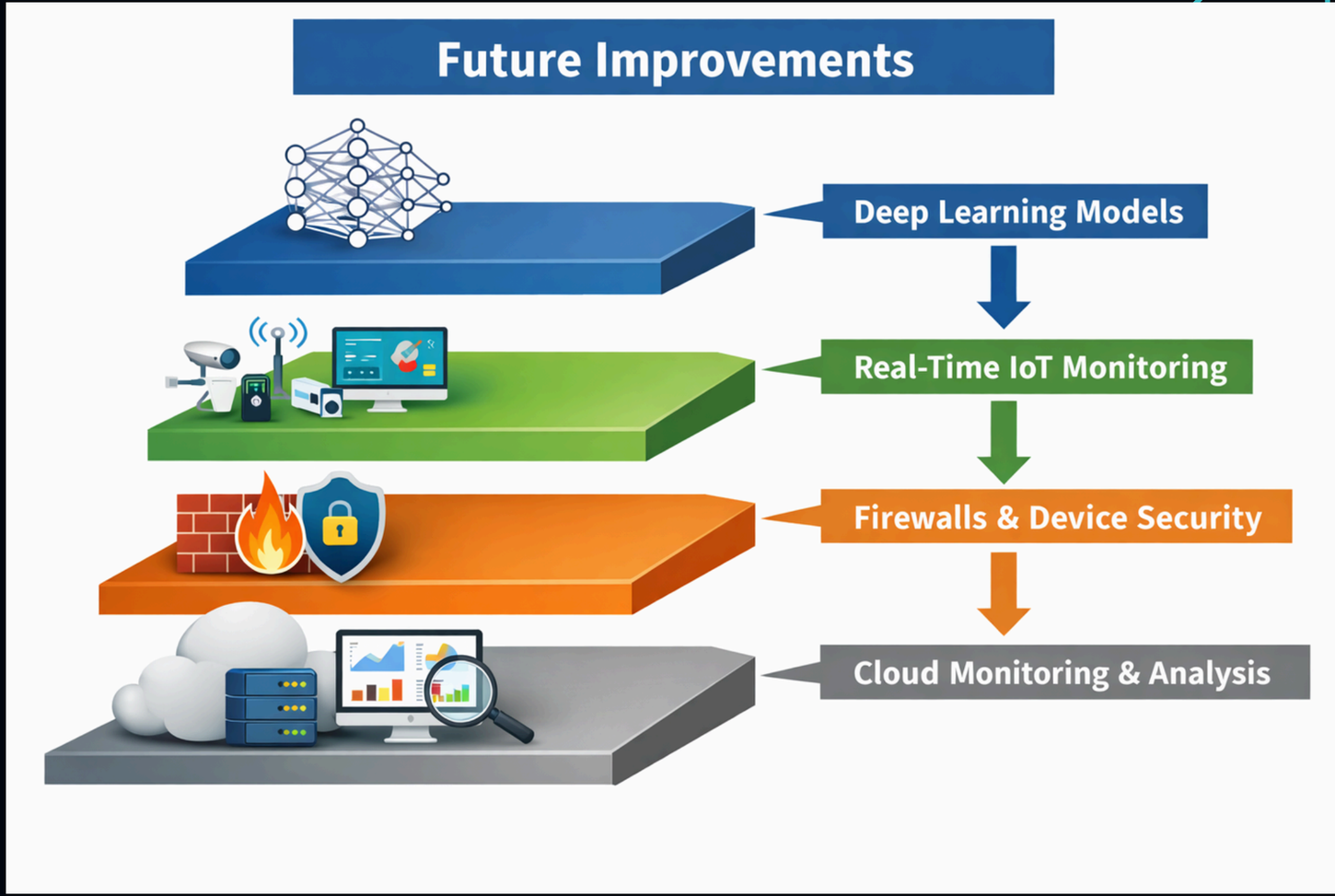
Malfunction of Critical Devices
Hacking of Vehicles, Cameras,
Medical Devices

Financial & Reputational Damage

Service Disruptions, Loss of Trust,
Economic Costs

FUTURE WORK

- Energy-aware IoT security
- Isolated IoT networks
- Defense-in-depth architecture
- Cloud-based detection and storage
- Alert prioritization by risk level



CONCLUSION

- IoT devices are highly vulnerable to botnet attacks
- Machine learning can effectively detect malicious IoT traffic
- Tree-based and ensemble models showed strong performance
- Binary detection is easier than family-level classification
- Efficient models are suitable for real-time IoT security