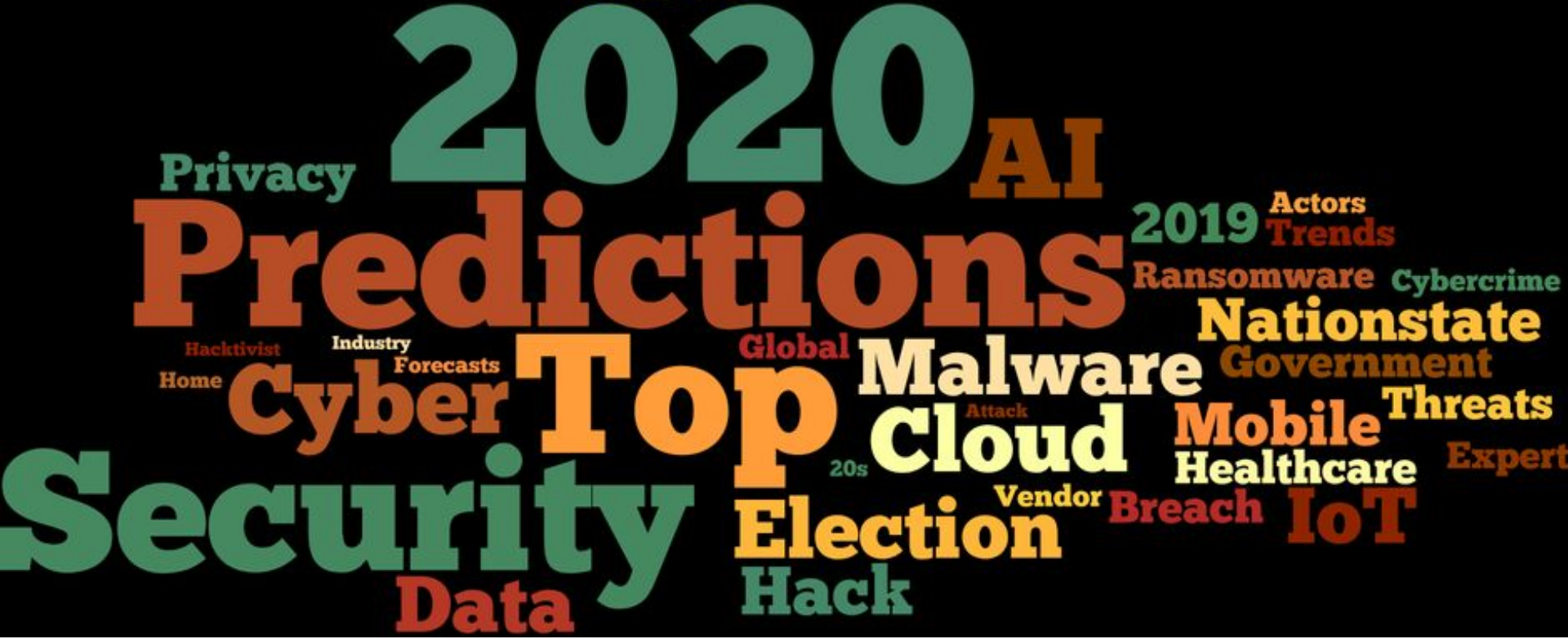



# Predicting IoT Malware Attacks

Prabhakar Rangarao

# Malwares Target IoT Devices Plan

**64 Billion IoT Devices (2025)**

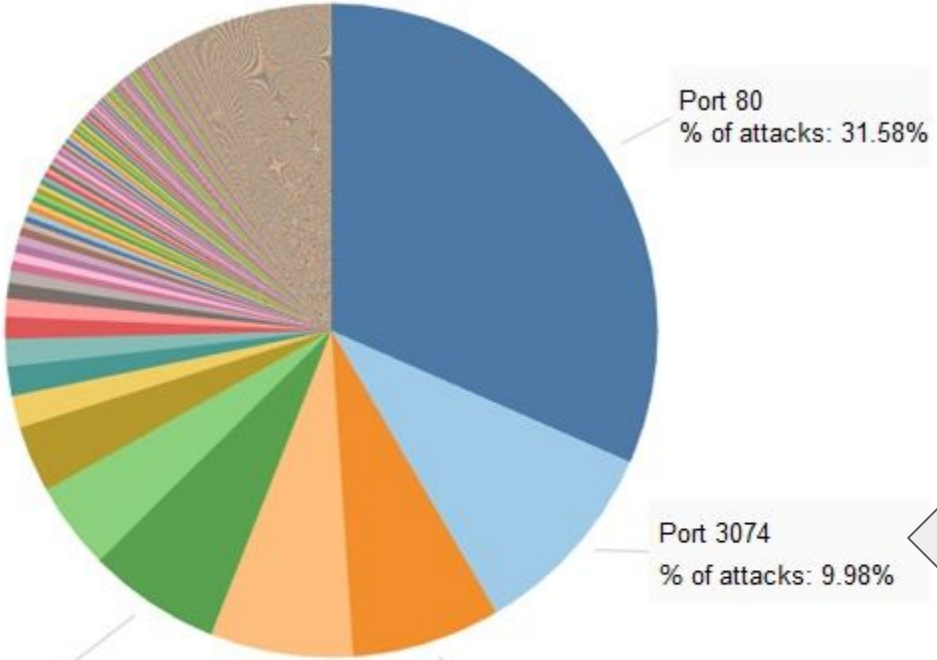
- **31 Billion** Internet of Things (IoT) connected by 2020
- **69%** of enterprises have more IoT devices than computer/network products
- **67%** of enterprises have experienced an IoT security incident
- **93%** of enterprises are planning to increase their spending to protect against attacks

**127** new IoT devices are connected to the web every second

# Malwares attack on common ports

Mirai


Bashlite





Ports used by Gamers

# Open Source makes it easy to launch botnet attacks

Provision PT737E
Security_Camera

Philips B120N10
Baby Monitor

Ecobee Thermostat

Provision PT838
Security_Camera


Ennino Doorbell

Samsung 1011N
Web Camera


Simple Home 1002
Security_Camera

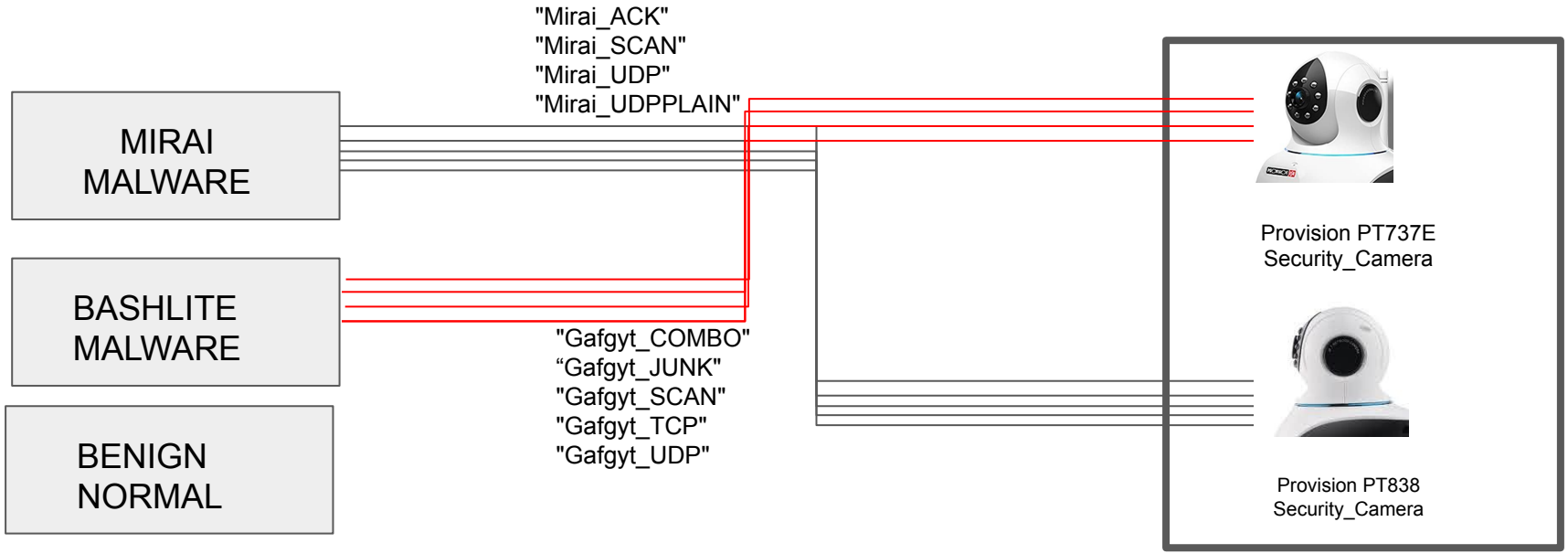
Simple Home 1003
Security_Camera

Damini Doorbell

# Goa.: Machine Learning Models as Mitigation Strategy

- Learning from the past attacks
  - 9 commercial IoT devices authentically infected by Mirai and BASHLITE (Source: UCI Machine Learning Repository)
- Sample datasets to train and test the models
  - Random Sampling with NO replacement from datasets and mix Benign, Mirai, and Bashlite traffic
- Classification Models for Prediction
  - Logistic Regression, KNN, Decision Tree, Random Forest, and XBBoost
- Score Model using Recall and Accuracy
  - Critical to minimize False Negative classification of malwares

# DATA UNDERSTANDING AND TRAFFIC TYPES



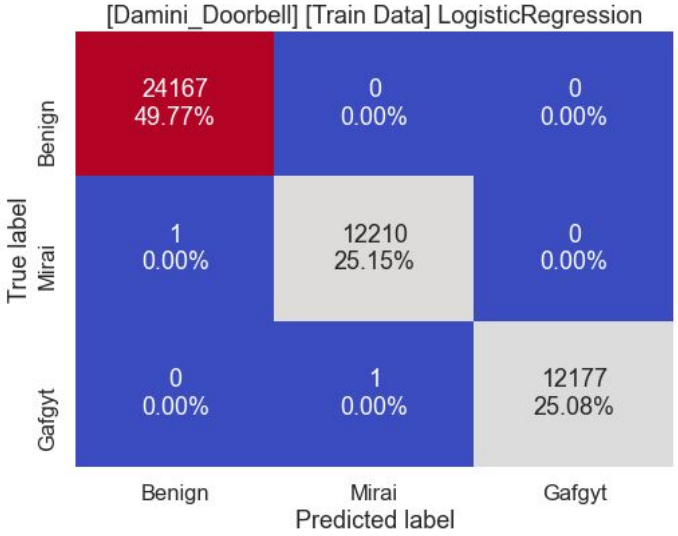

**7.5M**
**OBSERVATIONS**

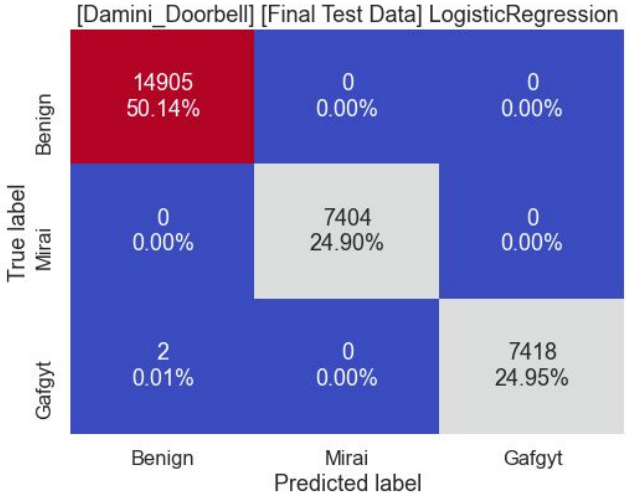
**115+(1)**
**FEATURES**

**9**
**IOT DEVICES**

# Model Comparison (Damini Doorbell)

## LOGISTIC REGRESSION

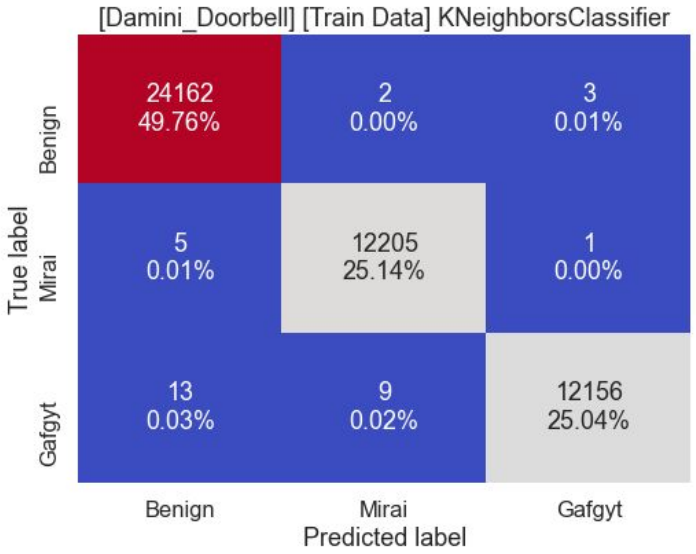


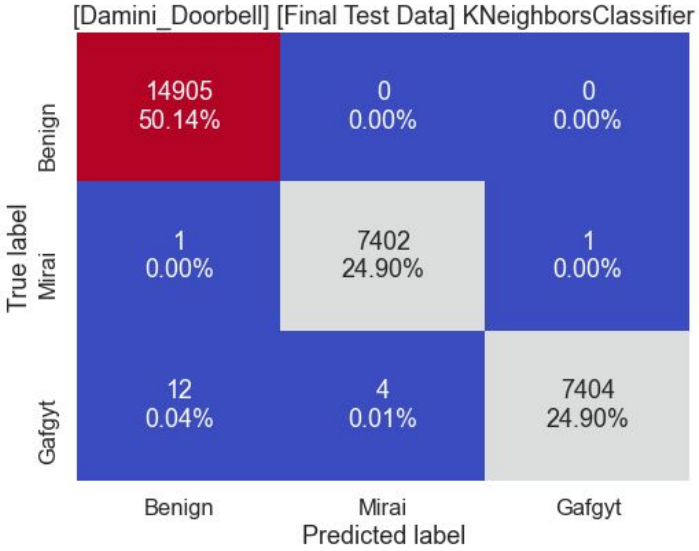




Training Recall 1.0
Final Test Recall 0.99

# Model Comparison (Damin Doorbell)

## KNN Model

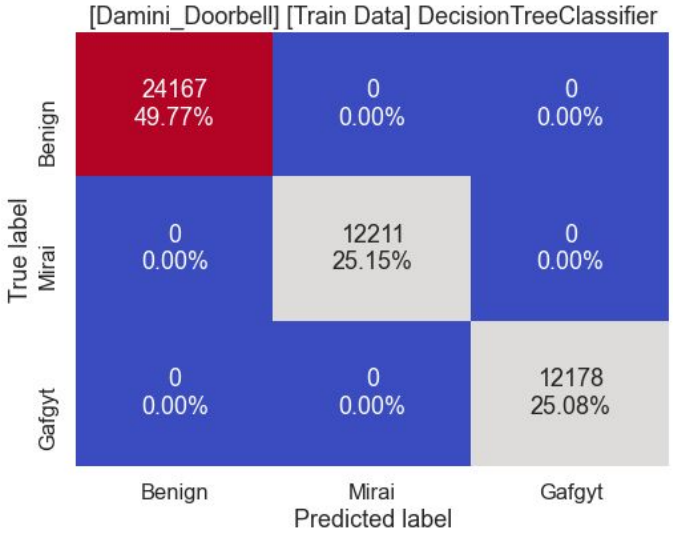


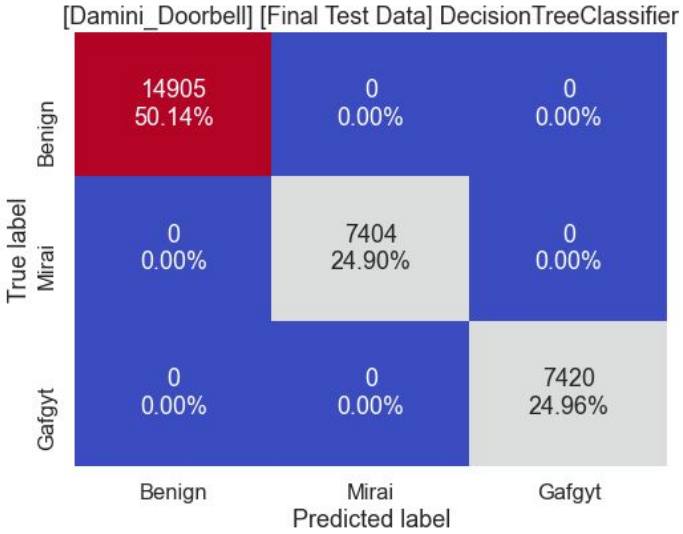




Train Data Recall Score 0.96
Test Data Recall Score 0.96

# Model Comparison (Damini Doorbell)

## Decision Tree Model

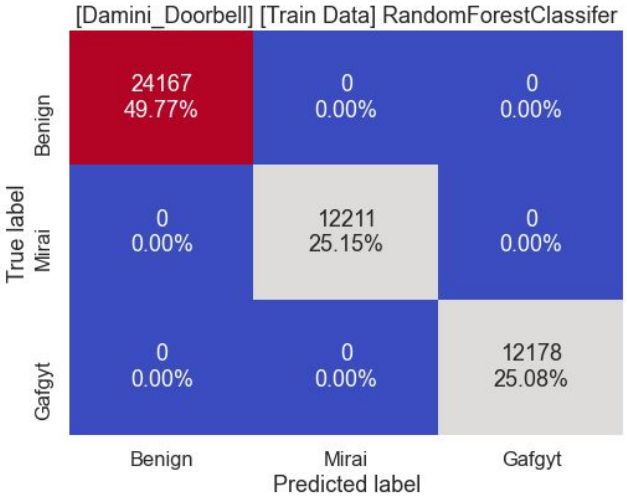


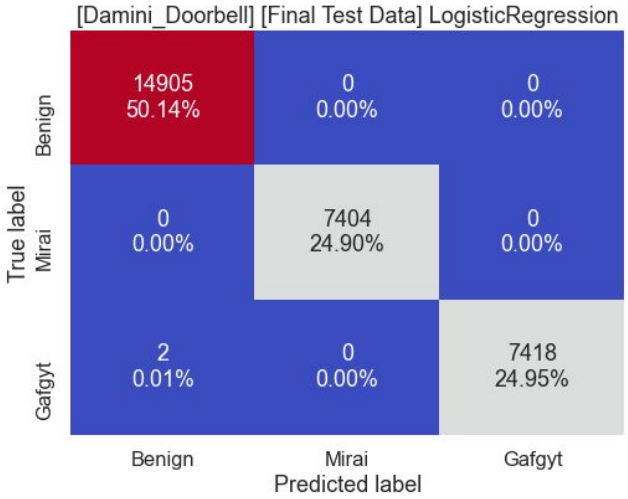




Train Data Recall Score 1.0
Test Data Recall Score 1.0

# Model Comparison (Damini Doorbell)

## Random Forest

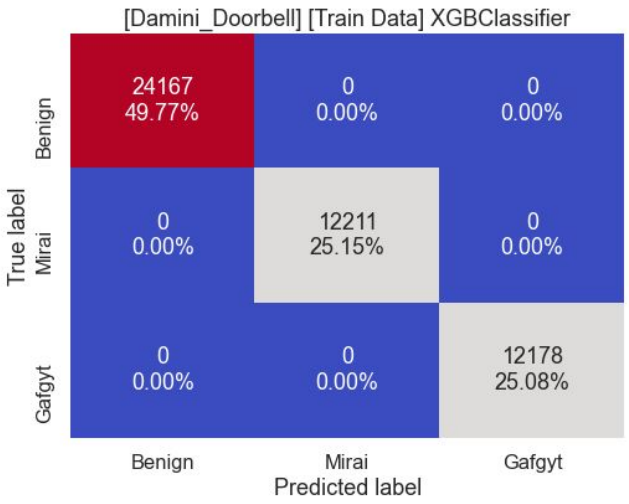


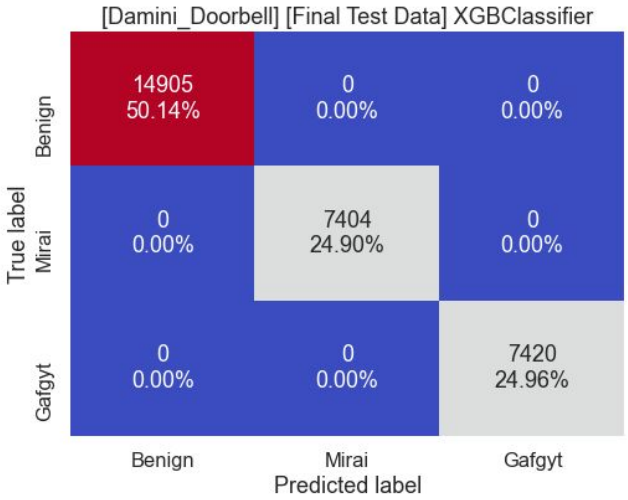




Train Data Recall Score 1.0
Test Data Recall Score 0.99

# Model Comparison (Damini Doorbell)

## XGBoost




Train Data Recall Score 1.0
Test Data Recall Score 1.0

# Model Summary

DECISION TREE

RANDOM FOREST

XGBOOST

3 models correctly predicted ALL malware attacks

XGBoost Model is the most promising decision-tree based ensemble ML Algorithm for detecting Malware Traffic

# Future Research Opportunities ..

- Expand to include additional data sources from latest botnet attacks
- Test it on AI TensorFlow/Keras Model
- Routers enabled by best Malware Anomaly Detection in real-time using Machine Learning Models in Real-time