
# Project Update

# Outline

1. Data Engineering Issues/Solutions (SMOTE)
2. BERT Model Performance
3. Framework Design
4. User Interface Design

INSyT

# Data Engineering

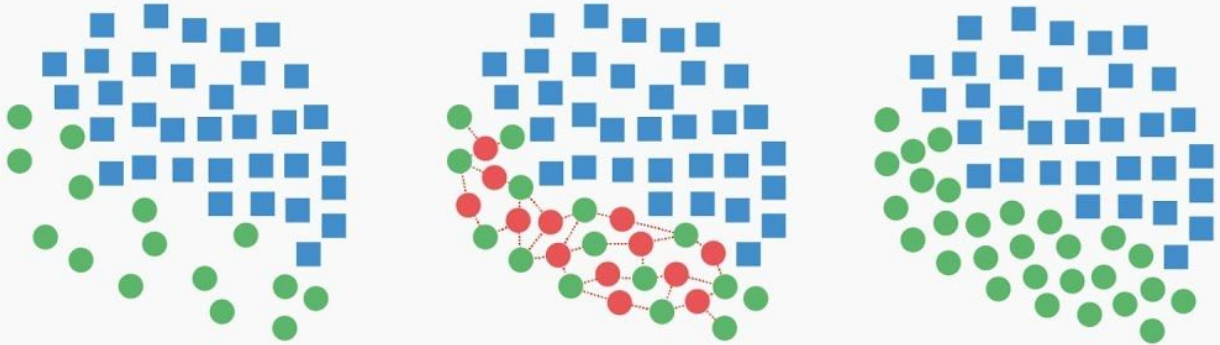
- Textual data -> Numeric representations
  - Embeddings
- Class imbalances
  - Re-groupings
- Sampling techniques
  - Duplicate sampling
  - SMOTE

```
label
benign                                                                  2423075
attacker_http:dirb:foothold                                             1689473
attacker:dnsteal:dnsteal-received                                        372856
attacker_http:foothold:wpscan                                             27671
dns_scan:foothold                                                          3305
foothold:service_scan                                                      1656
foothold:network_scan                                                       628
crack_passwords:escalate                                                    333
attacker_vpn:foothold                                                       224
attacker_http:foothold:webshell_cmd                                         126
escalate:webshell_cmd                                                       106
attacker_change_user:escalate                                                92
escalate:escalated_command:escalated_sudo_command                            91
foothold:wpscan                                                              64
attacker_http:foothold:service_scan                                          51
escalate:escalated_command:escalated_sudo_command:escalated_sudo_session     48
attacker:dnsteal:dnsteal-dropped                                             48
foothold:traceroute                                                          32
attacker_vpn:escalate                                                        28
attacker_http:foothold:webshell_upload                                       24
attacker:dnsteal:exfiltration-service                                        16
dirb:foothold                                                                12
attacker_change_user:escalate:escalated_command:escalated_sudo_command       8
Name: count, dtype: int64
```

# SMOTE

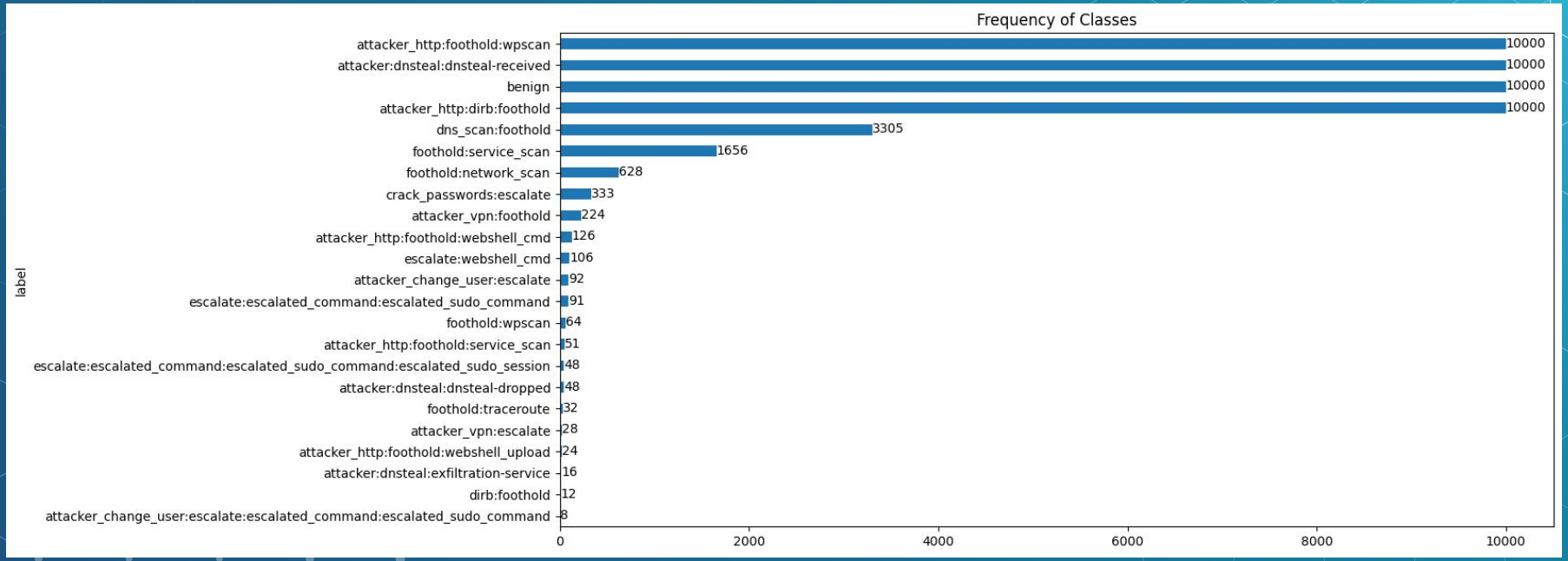
## HANDLE IMBALANCED DATASET

Synthetic Minority Oversampling Technique



Original Dataset | Generating Samples | Resampled Dataset

# BERT: Data

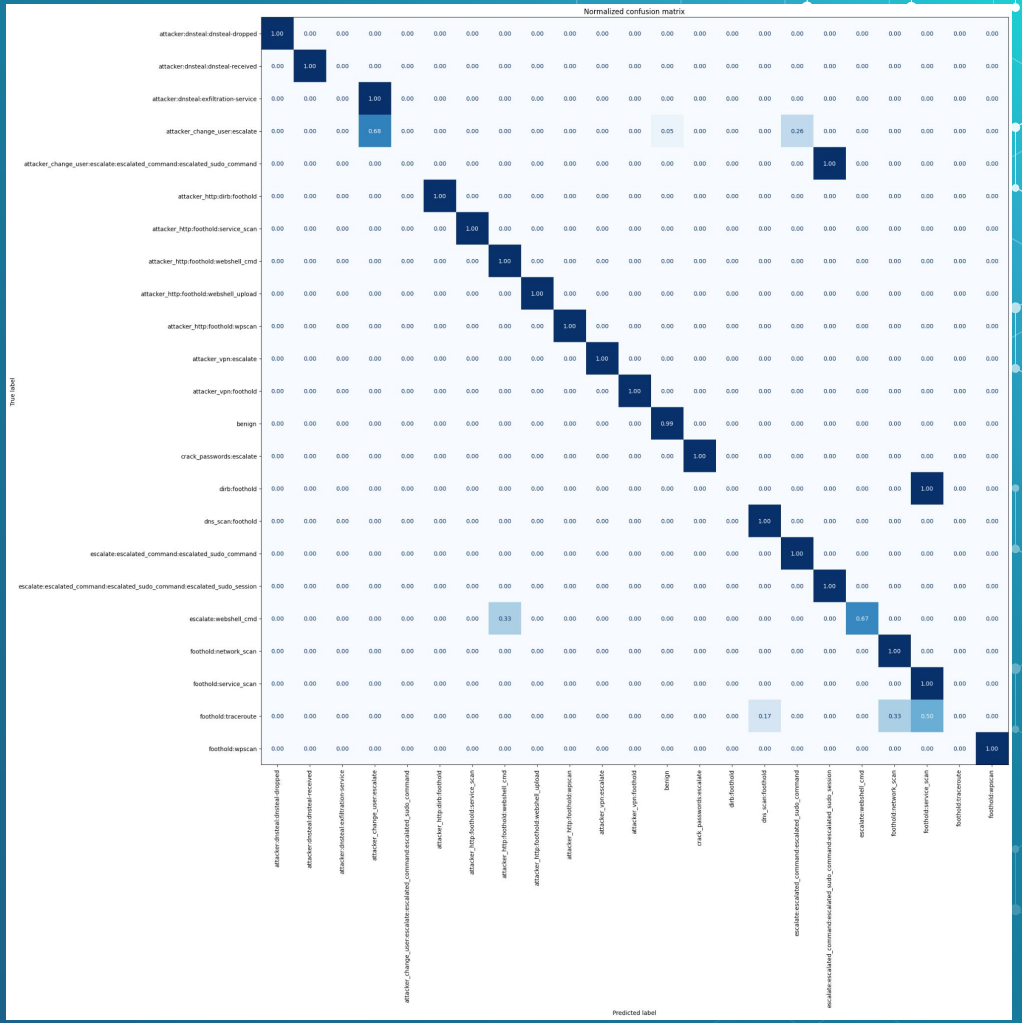

Frequency of Classes
label
attacker_http:foothold:wpscan
attacker:dnsteal:dnsteal-received
benign
attacker_http:dirb:foothold
dns_scan:foothold
foothold:service_scan
foothold:network_scan
crack_passwords:escalate
attacker_vpn:foothold
attacker_http:foothold:webshell_cmd
escalate:webshell_cmd
attacker_change_user:escalate
escalate:escalated_command:escalated_sudo_command
foothold:wpscan
attacker_http:foothold:service_scan
escalate:escalated_command:escalated_sudo_command:escalated_sudo_session
attacker:dnsteal:dnsteal-dropped
foothold:traceroute
attacker_vpn:escalate
attacker_http:foothold:webshell_upload
attacker:dnsteal:exfiltration-service
dirb:foothold
attacker_change_user:escalate:escalated_command:escalated_sudo_command
10000
10000
10000
10000
3305
1656
628
333
224
126
106
92
91
64
51
48
48
32
28
24
16
12
8
0
2000
4000
6000
8000
10000

# BERT: Results

| Epoch | Training Loss | Validation Loss | Accuracy | F1 |
|---|---|---|---|---|
| 1 | 0.285 | 0.071 | 0.981 | 0.979 |
| 2 | 0.053 | 0.030 | 0.993 | 0.992 |
| 3 | 0.029 | 0.024 | 0.995 | 0.994 |

| Test Loss | Test Accuracy | Test F1 |
|---|---|---|
| 0.0199 | 0.996 | 0.995 |

# BERT: Results




# Framework Design and Implementation

# Framework Design

- Goal: Set up as a PyPI Package for easy installation

# Framework Design

- Goal: Set up as a PyPI Package for easy installation

## Installation

```
pip install insyt

pip install git+https://github.com/Isaacwilliam4/Network_Intrusion.git
```

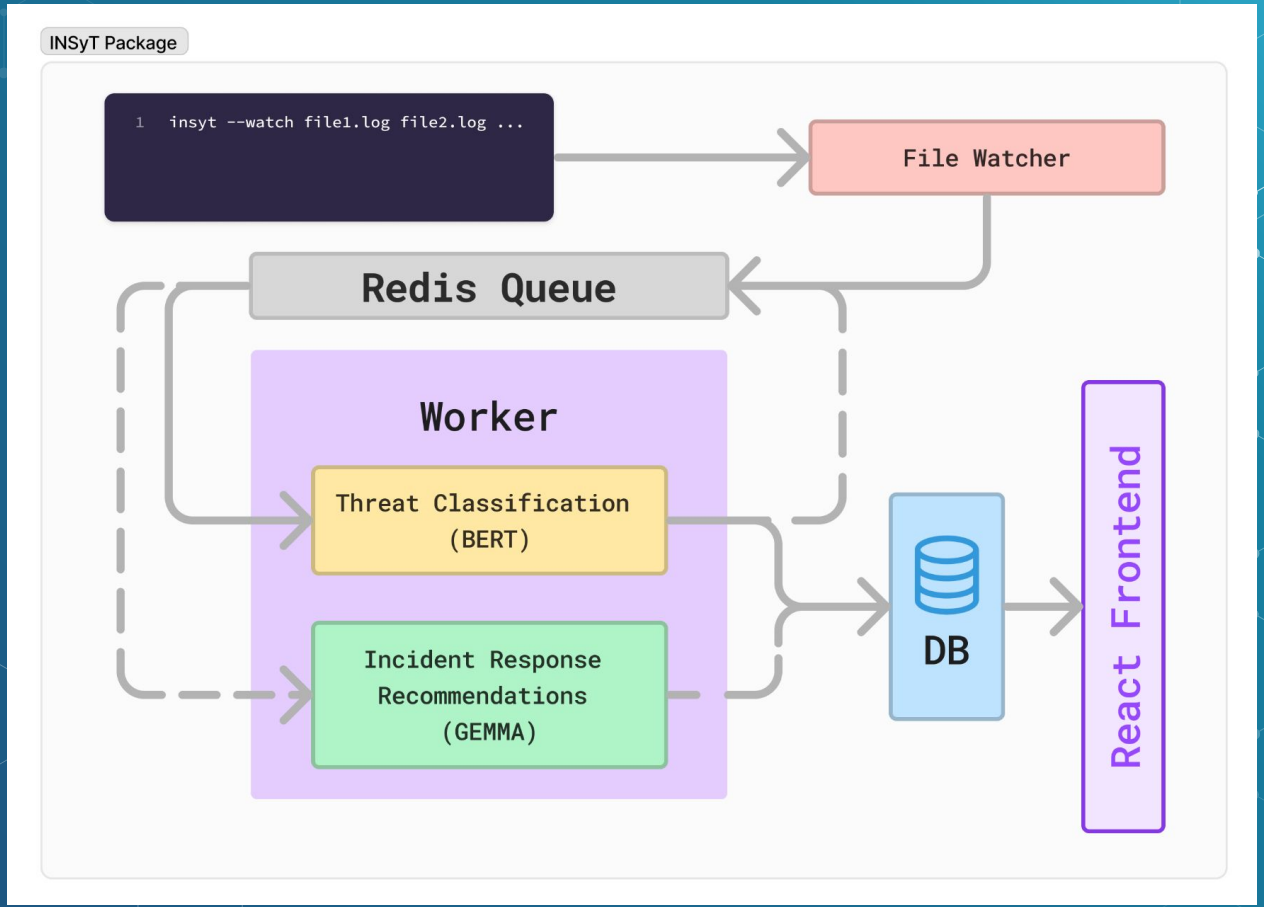
# setup.cfg

```
[metadata]
name = insyt
version = 0.0.1
license = Apache License 2.0
description = Innovative Network Security Technologies
long_description = file: README.md
author = Damon Tingey
author_email = damon.tingey@byu.edu

[options]
packages = find:
python_requires = >=3.10
install_requires =
    pandas
    watchdog
    tokenizers
    transformers
    torch
    ollama
    rq
[options.package_data]
insyt =
    py.typed
    **/Modelfile

[options.packages.find]
    exclude = insyt.tests
```

# Framework Design

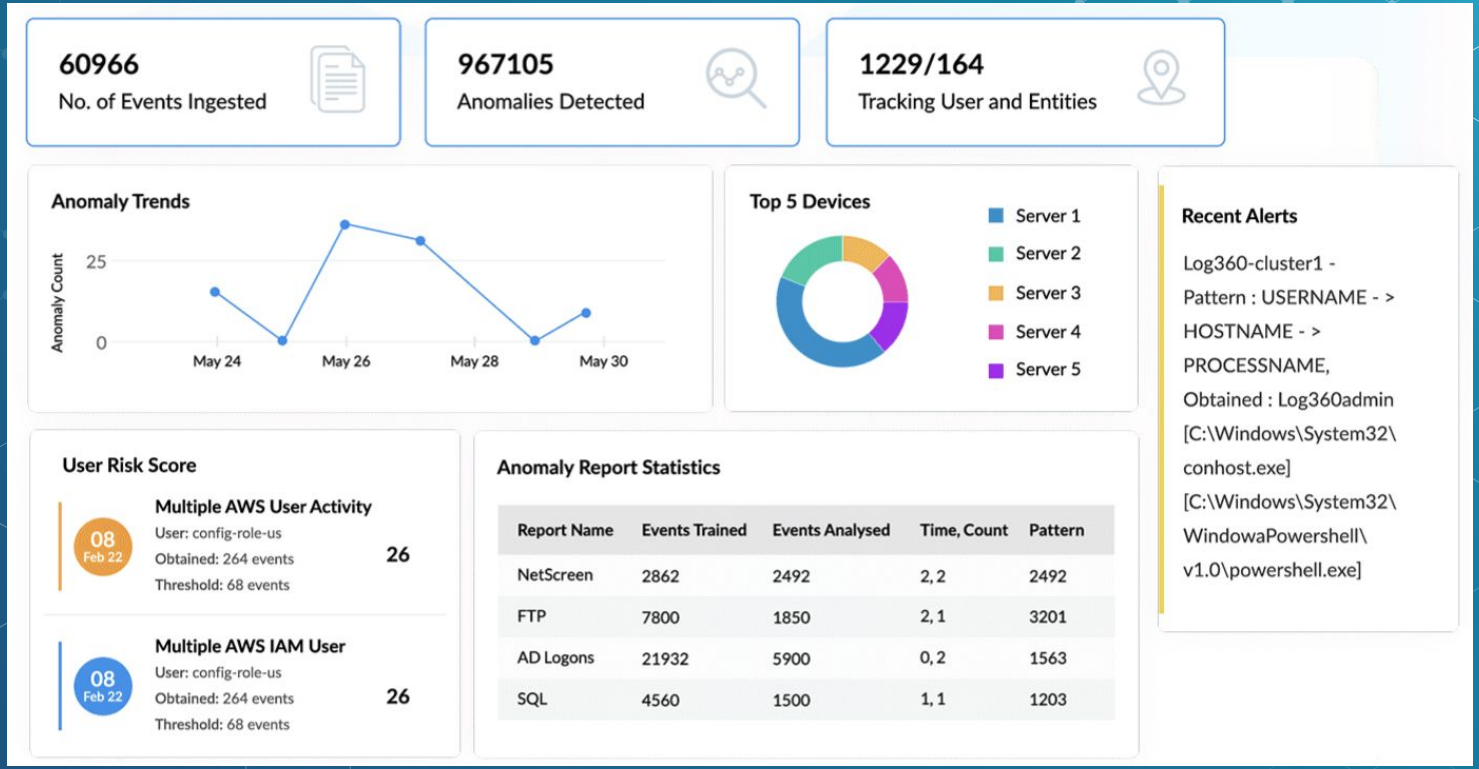



# Prototype Demo

# Framework Next Steps

- Set up background daemon
- Iron out edge cases (log lines added in various places, etc.)
- Develop a proper cli interface
- Hosting on PyPI

# User Interface Design



# User Interface Next Steps

- Set up separate pages for individual log
- Connect the real data to the graph library
- Implement the alarming system through Amazon SQS


# Questions?