

Configuration Manual

MSc Research Project
Msc. In Cybersecurity

Anushka Sharma
X23386592

School of Computing
National College of Ireland

Supervisor: Jawad Salahuddin

National College of Ireland
MSc Project Submission Sheet



School of Computing

ANUSHKA SHARMA

Student Name:

Student ID: X23386592

Programme: Msc. In Cybersecurity **Year:** 2025

Module: Practicum

Lecturer: Jawad Salahuddin

Submission Due Date: 11/12/2025

Project Title: Leveraging Threat Intelligence Feeds for Predictive Threat Modeling and Attack Simulation in Enterprise Networks

663 9

Word Count: **Page Count:**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ANUSHKA SHARMA

Date: 11/12/2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

X23386592

1 1. OVERVIEW

This manual provides a comprehensive setup guide for replicating the experimental environment used in the research project:

Leveraging Threat Intelligence Feeds for Predictive Threat Modeling and Attack Simulation in Enterprise Networks.

The system consists of five main components:

1. **Data Integration Layer** – Multi-source threat intelligence collection
2. **Feature Engineering Layer** – IOC extraction and temporal analysis
3. **ML Classification Layer** – Ensemble model training (Random Forest, Extra Trees, XGBoost)
4. **Attack Simulation Layer** – VirtualBox-based isolated attack lab
5. **Defense Mapping Layer** – MITRE ATT&CK-aligned SIEM/EDR rules generation

2. DATASET SETUP

2.1 Obtaining Threat Intelligence Feeds

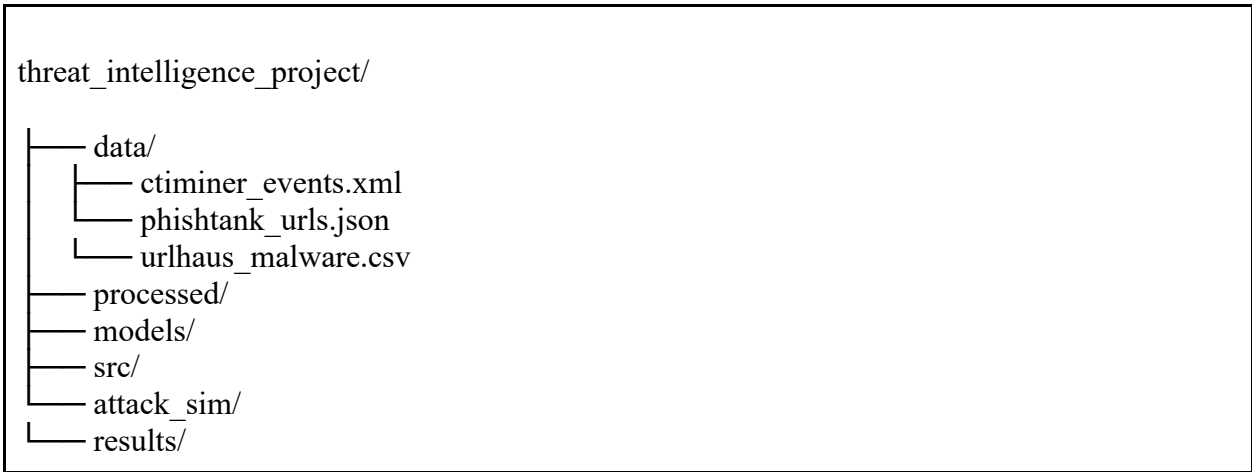
The system uses three open-source threat intelligence feeds:

Source	Format	Size	Access URL
CTIMiner	XML	11,116 events	[IEEE DataPort](https://dx.doi.org/10.21227/dpat-qd69)
PhishTank	JSON	10,000 URLs	[PhishTank](https://phishtank.org/)

URLhaus	CSV	10,000 URLs	[URLhaus](https://urlhaus.abuse.ch/)
---------	-----	-------------	--------------------------------------

2.2 Dataset Structure

Create the following directory structure:



2 2.3 Dataset Verification

Ensure each dataset contains the following record counts: Normal dataset: 495,000 rows

- CTIMiner: 11,116 events (2008–2019)
- PhishTank: 10,000 verified phishing URLs (2023–2025)
- URLhaus: 10,000 malware URLs (Nov 2025)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	3721446	#####	https://mu offline				malware_c_ClearFake	https://url anonymous											
2																			
3	3721445	#####	http://175. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
4																			
5	3721444	#####	http://222. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
6																			
7	3721443	#####	http://221. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
8																			
9	3721442	#####	http://39.8 online		#####	malware_c_32-bit,elf,r	https://url geenensp												
10																			
11	3721441	#####	https://f1n offline				malware_c_ClearFake	https://url anonymous											
12																			
13	3721440	#####	http://175. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
14																			
15	3721439	#####	http://222. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
16																			
17	3721437	#####	http://110. online		#####	malware_c_32-bit,elf,r	https://url geenensp												
18																			
19	3721438	#####	http://loq.l online		#####	malware_c_botnetdorm	https://url BlinkzSec												
20																			

Figure 1 Dataset Preview Showing Columns and Row Counts

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	phish_id	url	phish_det: submission	verified	verification	online	target												
2	9279673	https://lin	http://www	2025-11-3	yes	2025-11-3	yes	Other											
3	9279669	https://en	http://www	2025-11-3	yes	2025-11-3	yes	Other											
4	9279668	https://oz	http://www	2025-11-3	yes	2025-11-3	yes	Other											
5	9279667	https://li	http://www	2025-11-3	yes	2025-11-3	yes	Other											
6	9279666	https://na	http://www	2025-11-3	yes	2025-11-3	yes	Other											
7	9279665	https://fa	http://www	2025-11-3	yes	2025-11-3	yes	Other											
8	9279664	https://di	http://www	2025-11-3	yes	2025-11-3	yes	Other											
9	9279662	https://ca	http://www	2025-11-3	yes	2025-11-3	yes	Other											
10	9279661	https://ca	http://www	2025-11-3	yes	2025-11-3	yes	Other											
11	9279654	https://rs	http://www	2025-11-3	yes	2025-11-3	yes	Other											
12	9279653	https://li	http://www	2025-11-3	yes	2025-11-3	yes	Other											
13	9279651	https://su	http://www	2025-11-3	yes	2025-11-3	yes	Other											
14	9279649	https://mu	http://www	2025-11-3	yes	2025-11-3	yes	Other											
15	9279647	https://jo	http://www	2025-11-3	yes	2025-11-3	yes	Other											
16	9279646	https://li	http://www	2025-11-3	yes	2025-11-3	yes	Other											
17	9279645	https://gi	http://www	2025-11-3	yes	2025-11-3	yes	Other											
18	9279644	http://oi	http://www	2025-11-3	yes	2025-11-3	yes	Other											
19	9279643	https://li	http://www	2025-11-3	yes	2025-11-3	yes	Other											
20	9279642	https://id	http://www	2025-11-3	yes	2025-11-3	yes	Other											

Figure 2 Dataset Preview Showing Columns and Row Counts

	A	B	C	D	E	F	G	H	I	J	K	L
1	event_id	date	info	year	event_type	source_file	md5	sha1	sha256	ip_src	ip_dst	url
2		2008-10-01	Fritz_HOW-CHIN	2008	ReportEvent	CTIDataset_2008_ReportEvent.xml						politics.slashdot.
3	1	2008-11-11	556_10535_798	2008	ReportEvent	CTIDataset_2008_ReportEvent.xml						www.georgiapd.
4	3	2009-01-18	Ashmore_impac	2009	ReportEvent	CTIDataset_2009_ReportEvent.xml						prygi.blogspot.co
5	6	2009-03-29	ghostnet.pdf	2009	ReportEvent	CTIDataset_2009_ReportEvent.xml						www.internetwor
6	16	2008-04-08	253f59417c6c78	2010	MalwareEvent	CTIDataset_2010_253f59417c6c78	453f235fed7e22:b4e242e70f372b7fc846616f7ceabb9209875b47b3556cd7b0d0306c9					
7	17	2008-05-08	02677a0770268	2010	MalwareEvent	CTIDataset_2010_02677a0770268	384570c12c49bdfa16e57591b4bb004cc5f39fcb0f8dfed6dab7c4cb605587336bc71e					
8	18	2008-06-08	6e245522d710c	2010	MalwareEvent	CTIDataset_2010_6e245522d710c	3bd9022c36bcd7a5b3b8fcc83b80fdb225fe05902fe79751a6897101mcsmc.org mircr					
9	12	2008-06-08	69ef6009405232	2010	MalwareEvent	CTIDataset_2010_69ef6009405232	b1d8633f61b6e1cbcc28db5f06d22ae894439c6f0765f946f0f8e2dc0cmcsmc.org mircr					
10	23	2009-07-21	20ddc972f71c8e	2010	MalwareEvent	CTIDataset_2010_20ddc972f71c8e	3d76e5c230f7638d7235cd8fa83b1baf551ac725807b3b18126a9d1f	google.homeunib				
11	21	2009-07-20	a4a63756c39e3	2010	MalwareEvent	CTIDataset_2010_a4a63756c39e3	862b8a54edb4f37d508dac533c58f8a834e8842cf99607d526ae19c9058c26fdbe9b1f	google.homeunib				
12	19	2009-07-21	2794cacb3f177f	2010	MalwareEvent	CTIDataset_2010_2794cacb3f177f	9a708d2f80458a05b58c82010d0c0b226b48d5400aa74522596899c	google.homeunib				
13	14	2009-07-21	1326879b25dd0	2010	MalwareEvent	CTIDataset_2010_1326879b25dd0	c1ea57030a8721003157b395a67443c1dd0b332d25b2b49f80c6428	google.homeunib				
14	8	2009-07-21	2f6c8d68392839	2010	MalwareEvent	CTIDataset_2010_2f6c8d68392839	21293eb34d0bdi865adfc1f9c7ac74b049007edccb3539c574213516	google.homeunib				
15	26	2018-04-13	fc408bbd03ca2e	2010	MalwareEvent	CTIDataset_2010_b8a177d99854c069d86a9660d7fc408bbd03ca2ee5dd4d2694180cc5a779ef6e16ee	filoups.info					
16	11	2018-04-13	09b9b57a5fd05e	2010	MalwareEvent	CTIDataset_2010_d112a2ed6c675b2c04cb13f7dbf09b9b57a5fd0590c204610a1ed0f68ca9c8c5bab53	filoups.info goog					
17	10	2018-04-13	38db778cae20b	2010	MalwareEvent	CTIDataset_2010_273a51aada271f0522f923dde96i38db778cae20b9c8a16cd4a378bc4e08865b0098d	google-analitics.					

Figure 3 Dataset Preview Showing Columns and Row Counts

3. PYTHON ENVIRONMENT SETUP

3.1 Required Libraries

Create a file named requirements.txt with the following content:

```
scikit-learn==1.3.2
xgboost==2.0.2
pandas==2.1.4
numpy==1.24.3
lxml==4.9.3
requests==2.31.0
matplotlib==3.7.2
seaborn==0.12.2
joblib==1.3.2
```

3.2 Installation

Open terminal in the project directory and run:

```
pip install -r requirements.txt
```

4. MODEL TRAINING CONFIGURATION

4.1 Feature Engineering Settings

Configure in src/features.py:

```
IOC_FEATURES = [
    'has_md5', 'has_sha1', 'has_sha256', 'hash_count',
    'has_ip_src', 'has_ip_dst', 'ip_count',
    'has_url', 'url_count',
    'has_filename', 'filename_count',
    'has_cve', 'cve_count',
    'has_email',
    'month', 'day_of_week', 'quarter', 'is_weekend', 'days_since_start',
    'ioc_richness', 'is_malware_event'
]

THREAT_TYPES = [
    'APT', 'Botnet', 'Exploit', 'Malware',
    'Phishing', 'Ransomware', 'Trojan/RAT'
]
```

4.2 Model Hyperparameters

Configure model hyperparameters in src/config.py:

```
RANDOM_FOREST = {  
    'n_estimators': 100,  
    'max_depth': 20,  
    'class_weight': 'balanced',  
    'random_state': 42  
}
```

```
EXTRA_TREES = {  
    'n_estimators': 100,  
    'max_depth': 20,  
    'random_state': 42  
}
```

```
XGBOOST = {  
    'n_estimators': 100,  
    'max_depth': 6,  
    'learning_rate': 0.3,  
    'random_state': 42  
}
```

4.3 Training the Models

Run the training pipeline:

```
python src/model_training.py
```

This will generate four model files in the models/ directory:

- models/random_forest.pkl
- models/extra_trees.pkl
- models/xgboost_model.pkl

5. VIRTUAL MACHINE SETUP

5.1 Network Configuration

Create two VMs with **Host-Only Network** (192.168.56.0/24):

VM	Role	IP Address	Purpose
Ubuntu 22.04	Attacker	192.168.56.10	Attack Execution
Windows 10 Pro	Victim	192.168.56.20	Log Collection

Figure 7: VirtualBox Network Settings

5.2 Attacker VM Setup

Install tools:

```
sudo apt update
sudo apt install nmap netcat curl python3 python3-pip
```

Copy attack Scripts

```
scp attack_sim/execute_attacks.sh ubuntu@192.168.56.10:/home/ubuntu/
scp attack_sim/run_atomic_attacks.sh ubuntu@192.168.56.10:/home/ubuntu/
```

5.3 Victim VM Setup

Install Sysmon v15.15 and enable logging:

1. Download Sysmon from [Microsoft Sysinternals](#)
2. Run with config:

```
sysmon.exe -accepteula -i sysmonconfig-export.xml
```


3. Enable PowerShell logging:

```
Enable-PSLogging.ps1
```

6. FLASK LOG COLLECTION & VALIDATION SETUP

6.1 Log Collection Script

Copy collect_attack_logs.ps1 to Victim VM and run after attacks:

```
powershell -ExecutionPolicy Bypass -File collect_attack_logs.ps1
```

6.2 Log Output Location

Logs will be saved as CSVs in:

```
C:\AttackLogs\  
| security_events.csv  
| sysmon_events.csv  
| network_connections.csv  
| powershell_logs.csv
```

The application will be available at <http://localhost:5000>

6.3 Using the Web Application

Step 1: Upload a CSV file containing sensor data

Step 2: View trust score results and visualisations

7. DEFENSE MAPPING & SIEM/EDR CONFIGURATION

7.1 MITRE ATT&CK Mapping

Run the defense mapper:

```
python src/mitre_mapper.py
```

This generates:

- defense/prioritized_strategies.json
- defense/siem_rules/ (49 Splunk/Wazuh rules)
- defense/edr_configs/ (49 EDR configurations)

SIEM Rule Example (Splunk)

Example rule for Phishing detection:

```
spl
index=windows EventCode=4688
| search ProcessName="*powershell*" CommandLine="*Invoke-WebRequest*"
| eval technique="T1566.001"
| table _time, host, user, technique, CommandLine
```

RUNNING THE COMPLETE SYSTEM

Execution Steps

Start Victim logging (Windows VM):

```
powershell
```

```
Start-SysmonLogging
```

1. **Run attacks** (Ubuntu VM):

```
bash
```

```
./execute_attacks.sh
```

2. **Collect logs** (Windows VM):

```
powershell
```

```
./collect_attack_logs.ps1
```

3. **Validate model predictions:**

```
bash
```

```
python src/validate.py --logdir C:\AttackLogs\
```

4. **Generate defense reports:**

```
bash
```

```
python src/generate_defense_report.py
```

7.2 Expected Results

- **Model validation accuracy:** ~83.69% (Random Forest)
- **Attack detection accuracy:** ~14.25% (network-based validation)
- **Defense artifacts:** 49 SIEM rules + 49 EDR configs
- **Log summary:** ~252,866 Windows events

TimeCreated																
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	SourceIp:192.168.1.100 SourcePort:555 DestinationIp:192.168.1.101 DestinationPort:443 DestinationHost:192.168.1.101	443	tcp	Phishing	0.5419952818	Phishing	0.4505534157	Phishing	0.3549503353	Phishing	0.79078203	2025-12-02 11:0	2025-12-02 11:0			
	Network:correl RuleName: UtcTime: 2025-1 ProcessId: 50 ProcessName: 1120 Image: C:\Program Files\Internet User: NT AUTHORITY\SYSTEM Protocol: tcp Initiated: true SourceIp: 192.168.1.100 SourcePort: 555 DestinationIp: 192.168.1.101 DestinationPort: 443 DestinationHost: 192.168.1.101															
2	SourceIp:192.168.1.100 SourcePort:555 DestinationIp:192.168.1.101 DestinationPort:443 DestinationHost:192.168.1.101	443	tcp	Phishing	0.5419952818	Phishing	0.4505534157	Phishing	0.3549503353	Phishing	0.79078203	2025-12-02 11:0	2025-12-02 11:0			
	Network:correl RuleName: UtcTime: 2025-1 ProcessId: 50 ProcessName: 1120 Image: C:\Program Files\Internet User: NT AUTHORITY\SYSTEM Protocol: tcp Initiated: true SourceIp: 192.168.1.100 SourcePort: 555 DestinationIp: 192.168.1.101 DestinationPort: 443 DestinationHost: 192.168.1.101															
3	SourceIp:192.168.1.100 SourcePort:555 DestinationIp:192.168.1.101 DestinationPort:443 DestinationHost:192.168.1.101	443	tcp	Phishing	0.5419952818	Phishing	0.4505534157	Phishing	0.3549503353	Phishing	0.79078203	2025-12-02 11:0	2025-12-02 11:0			
	Network:correl RuleName: UtcTime: 2025-1 ProcessId: 50 ProcessName: 1120 Image: C:\Program Files\Internet User: NT AUTHORITY\SYSTEM Protocol: tcp Initiated: true SourceIp: 192.168.1.100 SourcePort: 555 DestinationIp: 192.168.1.101 DestinationPort: 443 DestinationHost: 192.168.1.101															
4	SourceIp:192.168.1.100 SourcePort:555 DestinationIp:192.168.1.101 DestinationPort:443 DestinationHost:192.168.1.101	443	tcp	Phishing	0.5419952818	Phishing	0.4505534157	Phishing	0.3549503353	Phishing	0.79078203	2025-12-02 11:0	2025-12-02 11:0			
	Network:correl RuleName: UtcTime: 2025-1 ProcessId: 50 ProcessName: 1120 Image: C:\Program Files\Internet User: NT AUTHORITY\SYSTEM Protocol: tcp Initiated: true SourceIp: 192.168.1.100 SourcePort: 555 DestinationIp: 192.168.1.101 DestinationPort: 443 DestinationHost: 192.168.1.101															
5	SourceIp:192.168.1.100 SourcePort:555 DestinationIp:192.168.1.101 DestinationPort:443 DestinationHost:192.168.1.101	443	tcp	Phishing	0.5419952818	Phishing	0.4505534157	Phishing	0.3549503353	Phishing	0.79078203	2025-12-02 11:0	2025-12-02 11:0			
	Network:correl RuleName: UtcTime: 2025-1 ProcessId: 50 ProcessName: 1120 Image: C:\Program Files\Internet User: NT AUTHORITY\SYSTEM Protocol: tcp Initiated: true SourceIp: 192.168.1.100 SourcePort: 555 DestinationIp: 192.168.1.101 DestinationPort: 443 DestinationHost: 192.168.1.101															

Figure 4 Results.csv