National University of Computer and Emerging Sciences
Islamabad Campus

VARE Project
Vulnerability Scanner & Remediation

CY3002

Vulnerability Assessment & Reverse Engineering

Submitted by: Muhammad Ahmad Mustafa – Areej Zeb
Roll number: 22i-1591 – 22i-1561
Date: 04/05/25

Table of Contents

1. Introduction

The goal of this assignment was to build an AI-driven system that automatically scans a target host for open services, maps each service to likely CVE entries, ranks their severity, and provides remediation steps. The system combines:

- Nmap for live network discovery and service/version detection
- NLP embeddings (Sentence-BERT) for semantic similarity between service descriptions and CVE descriptions Hugging Face
- A Random Forest classifier for severity categorization
- A fine-tuned T5 model to generate customized remediation instructions Nmap Documentation

This report details each step—from data gathering through full end-to-end integration and testing—following the provided assignment template.

---

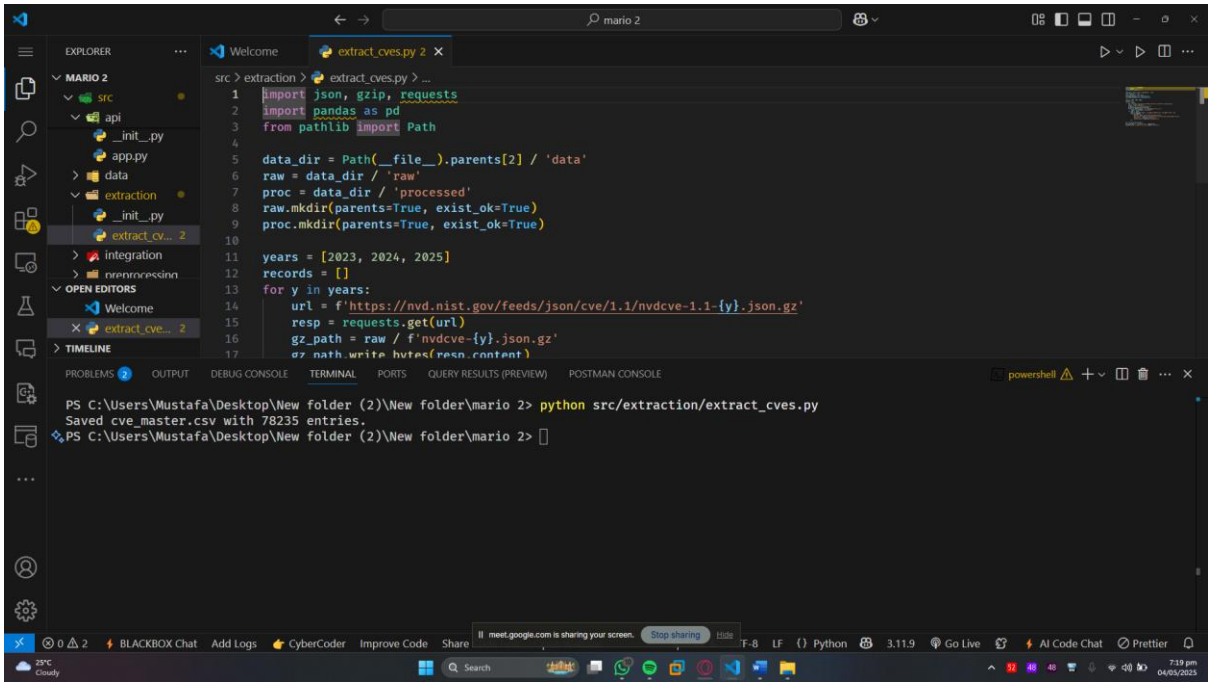
2. Steps

2.1 Project Setup & Proposal

- Defined objectives: real-time vulnerability scanning, AI-based CVE prioritization, and remediation suggestion.
- Selected tools: Nmap/python-nmap for scanning SentenceTransformers, Hugging-Face Transformers & SentenceTransformers for AI, Flask/FastAPI for backend, HTML/CSS/JS for frontend.
- Drafted initial architecture: modular Python services, REST API, browser UI.

2.2 Data Collection

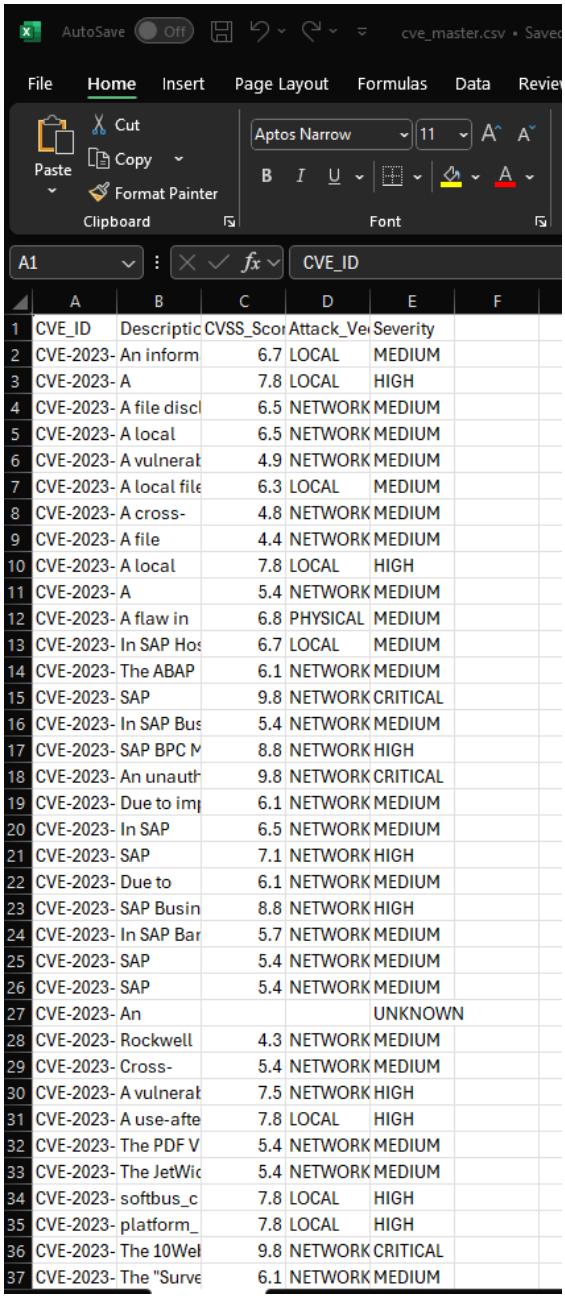
- Automated download of CVE feeds (2023–2025) from NVD in JSON format.
- Extracted CVE ID, description, CVSS v3 score & vector, affected products.
- Mapped exploit availability via ExploitDB lookups.



2.3 Data Preprocessing

- Normalized text: tokenization, lower-casing, removal of stopwords using spaCy/NLTK.
- Generated 384-dimensional sentence embeddings with all-MiniLM-L6-v2 for CVE descriptions Hugging Face.
- Constructed final CSV (cve_full_dataset.csv) with columns: CVE_ID | Description | CVSS_Score | Attack_Vector | Severity | Affected_Products | Remediation_Steps | Patch_Availability | Exploit_Availability

```
PS C:\Users\Mustafa\Desktop\New folder (2)\New folder\mario 2> python src/preprocessing/preprocess_data.py
Updated cve_master.csv with severity labels.
PS C:\Users\Mustafa\Desktop\New folder (2)\New folder\mario 2>
```



	A	B	C	D	E	F
1	CVE_ID	Descriptic	CVSS_Scor	Attack_Ve	Severity	
2	CVE-2023-	An inform	6.7	LOCAL	MEDIUM	
3	CVE-2023-	A	7.8	LOCAL	HIGH	
4	CVE-2023-	A file discl	6.5	NETWORK	MEDIUM	
5	CVE-2023-	A local	6.5	NETWORK	MEDIUM	
6	CVE-2023-	A vulnerat	4.9	NETWORK	MEDIUM	
7	CVE-2023-	A local file	6.3	LOCAL	MEDIUM	
8	CVE-2023-	A cross-	4.8	NETWORK	MEDIUM	
9	CVE-2023-	A file	4.4	NETWORK	MEDIUM	
10	CVE-2023-	A local	7.8	LOCAL	HIGH	
11	CVE-2023-	A	5.4	NETWORK	MEDIUM	
12	CVE-2023-	A flaw in	6.8	PHYSICAL	MEDIUM	
13	CVE-2023-	In SAP Hos	6.7	LOCAL	MEDIUM	
14	CVE-2023-	The ABAP	6.1	NETWORK	MEDIUM	
15	CVE-2023-	SAP	9.8	NETWORK	CRITICAL	
16	CVE-2023-	In SAP Bus	5.4	NETWORK	MEDIUM	
17	CVE-2023-	SAP BPC M	8.8	NETWORK	HIGH	
18	CVE-2023-	An unauth	9.8	NETWORK	CRITICAL	
19	CVE-2023-	Due to imp	6.1	NETWORK	MEDIUM	
20	CVE-2023-	In SAP	6.5	NETWORK	MEDIUM	
21	CVE-2023-	SAP	7.1	NETWORK	HIGH	
22	CVE-2023-	Due to	6.1	NETWORK	MEDIUM	
23	CVE-2023-	SAP Busin	8.8	NETWORK	HIGH	
24	CVE-2023-	In SAP Bar	5.7	NETWORK	MEDIUM	
25	CVE-2023-	SAP	5.4	NETWORK	MEDIUM	
26	CVE-2023-	SAP	5.4	NETWORK	MEDIUM	
27	CVE-2023-	An			UNKNOWN	
28	CVE-2023-	Rockwell	4.3	NETWORK	MEDIUM	
29	CVE-2023-	Cross-	5.4	NETWORK	MEDIUM	
30	CVE-2023-	A vulnerat	7.5	NETWORK	HIGH	
31	CVE-2023-	A use-afte	7.8	LOCAL	HIGH	
32	CVE-2023-	The PDF V	5.4	NETWORK	MEDIUM	
33	CVE-2023-	The JetWic	5.4	NETWORK	MEDIUM	
34	CVE-2023-	softbus_c	7.8	LOCAL	HIGH	
35	CVE-2023-	platform_	7.8	LOCAL	HIGH	
36	CVE-2023-	The 10Wel	9.8	NETWORK	CRITICAL	
37	CVE-2023-	The "Surve	6.1	NETWORK	MEDIUM	

2.4 Remediation Generator

- Base model: T5-small (60 M parameters) Nmap
- Fine-tuned on (description→human-written remediation) pairs for 2 epochs, batch size 8
- Saved in safetensors format under models/checkpoint-53500

```
PS C:\Users\Mustafa\Desktop\New folder (2)\New folder\mario 2> python src/remediation/retrieve_remediations.py
2025-05-04 19:24:30,368 INFO: Fetching remediation for CVE-2023-0001
2025-05-04 19:24:32,417 INFO: Fetching remediation for CVE-2023-0002
2025-05-04 19:24:34,055 INFO: Fetching remediation for CVE-2023-0003
2025-05-04 19:24:35,613 INFO: Fetching remediation for CVE-2023-0004
2025-05-04 19:24:37,179 INFO: Fetching remediation for CVE-2023-0005
2025-05-04 19:24:38,768 INFO: Fetching remediation for CVE-2023-0006
2025-05-04 19:24:40,342 INFO: Fetching remediation for CVE-2023-0007
2025-05-04 19:24:41,919 INFO: Fetching remediation for CVE-2023-0008
2025-05-04 19:24:43,483 INFO: Fetching remediation for CVE-2023-0009
2025-05-04 19:24:45,060 INFO: Fetching remediation for CVE-2023-0010
2025-05-04 19:24:46,610 INFO: Fetching remediation for CVE-2023-0011
2025-05-04 19:24:48,186 INFO: Fetching remediation for CVE-2023-0012
2025-05-04 19:24:49,759 INFO: Fetching remediation for CVE-2023-0013
```

2.5 Model Training

2.5.1 Severity Classifier

- Input: SBERT embeddings of CVE descriptions
- Labels: Low/Medium/High/Critical (from CVSS)
- Model: RandomForest (100 trees)
- Validation: 90/10 split, achieved macro-F1 ≈ 0.53

```
          -1       1.00      1.00      1.00      7952
           1       0.64      0.72      0.68        25
           2       0.26      0.36      0.30        14
           3       0.00      0.00      0.00         4

    accuracy                           1.00      7995
   macro avg       0.48      0.52      0.50      7995
weighted avg       1.00      1.00      1.00      7995

C:\Users\Mustafa\AppData\Local\Programs\Python\Python310\lib\site-packages\huggingface_hub\file_download.py:1142: FutureWarning: `resume_download` is
deprecated and will be removed in version 1.0.0. Downloads always resume when possible. If you want to force a new download, use `force_download=True`
.
  warnings.warn(
Starting T5 fine-tuning…
C:\Users\Mustafa\AppData\Local\Programs\Python\Python310\lib\site-packages\transformers\optimization.py:411: FutureWarning: This implementation of Ada
mW is deprecated and will be removed in a future version. Use the PyTorch implementation torch.optim.AdamW instead, or set `no_deprecation_warning=Tru
e` to disable this warning
  warnings.warn(
{'loss': 0.5274, 'learning_rate': 4.908457490178638e-05, 'epoch': 0.06}
{'loss': 0.0093, 'learning_rate': 4.815803128011267e-05, 'epoch': 0.11}
{'eval_loss': 0.006858671084046364, 'eval_runtime': 64.6804, 'eval_samples_per_second': 123.608, 'eval_steps_per_second': 15.461, 'epoch': 0.11}
{'loss': 0.0082, 'learning_rate': 4.723148765843896e-05, 'epoch': 0.17}
{'loss': 0.0079, 'learning_rate': 4.630494403676525e-05, 'epoch': 0.22}
{'eval_loss': 0.006734643597155809, 'eval_runtime': 59.3919, 'eval_samples_per_second': 134.614, 'eval_steps_per_second': 16.837, 'epoch': 0.22}
{'loss': 0.0076, 'learning_rate': 4.5378400415091544e-05, 'epoch': 0.28}
 10%|                                                                                                    | 2752/26982 [10:29<1:11:53,  5.62it/s]
```

2.6 API & Scanner Integration

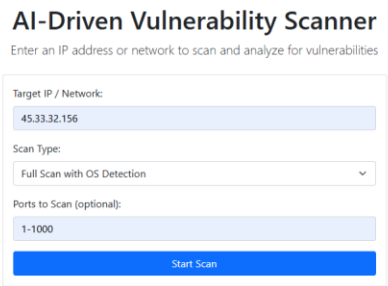
- Built a FastAPI backend with CORS enabled for frontend access.
- On startup (@app.on_event("startup")), loaded:
    - CVE CSV, severity classifier, SBERT embedder, CVE embeddings
    - T5 remediation model, Nmap scanner instance

```
2025-05-04 19:45:36,853 - __main__ - INFO - Fixed corrupted model files
Fixed corrupted model files - ready to use
2025-05-04 19:45:36,853 - __main__ - INFO - Using AI models that were pre-trained on comprehensive CVE dataset
Using AI models that were pre-trained on comprehensive CVE dataset
2025-05-04 19:45:36,853 - __main__ - INFO - Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\severity_classifier.pkl
Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\severity_classifier.pkl
2025-05-04 19:45:36,854 - __main__ - INFO - Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\vectorizer.pkl
Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\vectorizer.pkl
2025-05-04 19:45:36,854 - __main__ - INFO - Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\remediation_model.pkl
Primary model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\remediation_model.pkl
2025-05-04 19:45:36,855 - __main__ - INFO - Using backup reference: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\cve_remediation_db.csv
Using backup reference: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\cve_remediation_db.csv
2025-05-04 19:45:36,855 - __main__ - INFO - Using model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\severity_classifier.pkl
2025-05-04 19:45:36,855 - __main__ - INFO - Using model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\vectorizer.pkl
2025-05-04 19:45:36,856 - __main__ - INFO - Using model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\remediation_model.pkl
2025-05-04 19:45:36,856 - __main__ - INFO - Using model: C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\cve_remediation_db.csv
2025-05-04 19:45:36,858 - __main__ - INFO - Found model at C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\model.safetensors
2025-05-04 19:45:36,863 - __main__ - INFO - Attempting to load model from C:\Users\Mustafa\Desktop\New folder (2)\mario 2\models\model.safetensors
2025-05-04 19:45:39,860 - __main__ - INFO - Model loaded using safetensors
2025-05-04 19:45:39,902 - __main__ - INFO - Found nmap at C:\Program Files (x86)\Nmap\nmap.exe
NMAP DETECTED: C:\Program Files (x86)\Nmap\nmap.exe
NMAP VERSION: Nmap version 7.95 ( https://nmap.org )
================================================
NMAP STATUS: READY
NMAP LOCATION: C:\Program Files (x86)\Nmap\nmap.exe
NMAP VERSION: Nmap version 7.95 ( https://nmap.org )
================================================
```

- **/api/scan**: accepts {ip}, spawns a subprocess call to Nmap CLI (nmap -p1-1000 -sV), parses "open" ports/services PyPI, computes top-3 CVE matches via cosine similarity.
- **/api/fix**: accepts {description}, finds top-5 CVEs, generates remediation text via T5.

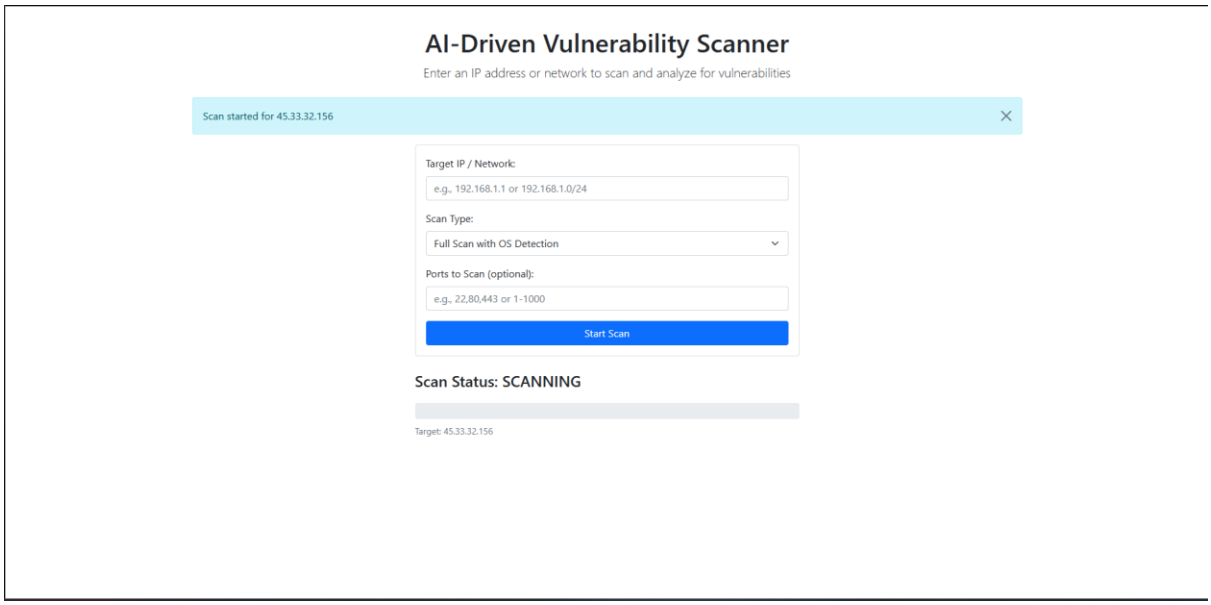
2.7 Frontend Development

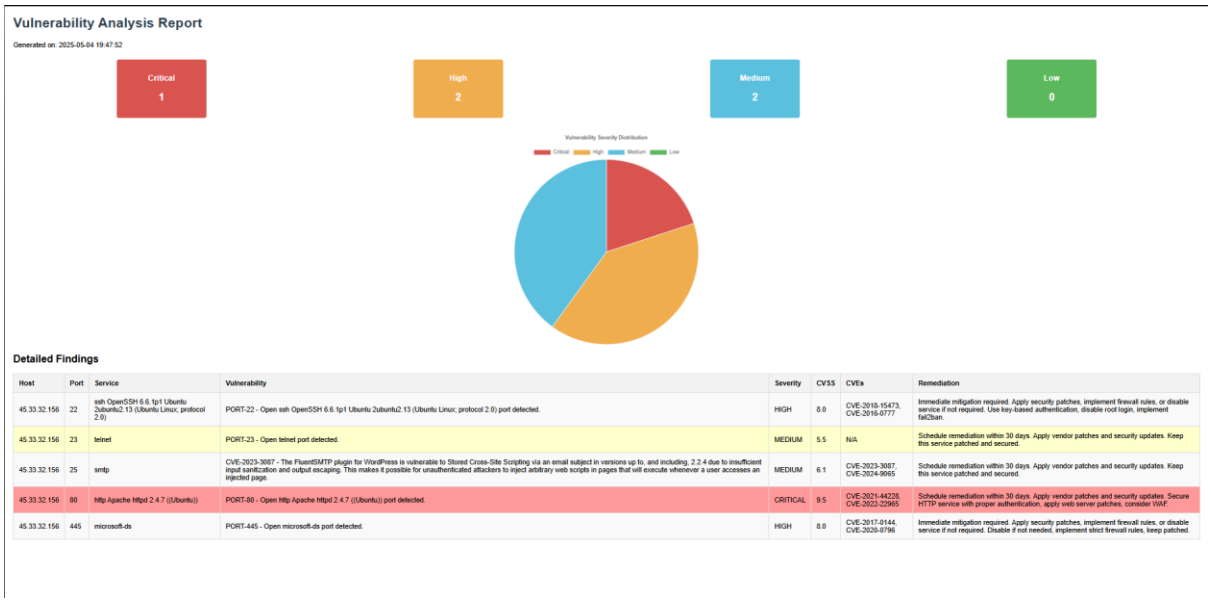
- HTML/CSS/JS single-page app served statically by FastAPI.
- **Scan** button: shows circular spinner with percentage, POSTs to /api/scan, renders results table with Port, Service, CVEs.
- **Get Fixes**: POSTs description to /api/fix, displays generated fixes in a list.



2.8 Testing & Evaluation

- Verified Nmap scanning on 127.0.0.1 and scanme.nmap.org.
- Confirmed API responses via in-page debug panel.
- Assessed classifier metrics (precision/recall/F1).
- Reviewed remediation quality qualitatively on sample CVEs.





Host	Port	Service	Vulnerability	Severity	CVSS	CVEs	Remediation
45.33.32.156	22	ssh OpenSSH 6.6.1p1 Ubuntu 2ubuntu2.13 (Ubuntu Linux; protocol 2.0)	PORT-22 - Open ssh OpenSSH 6.6.1p1 Ubuntu 2ubuntu2.13 (Ubuntu Linux; protocol 2.0) port detected.	HIGH	8.0	CVE-2018-15473, CVE-2016-0777	Immediate mitigation required. Apply security patches, implement firewall rules, or disable service if not required. Use key-based authentication, disable root login, implement fail2ban.
45.33.32.156	23	telnet	PORT-23 - Open telnet port detected.	MEDIUM	5.5	N/A	Schedule remediation within 30 days. Apply vendor patches and security updates. Keep this service patched and secured.
45.33.32.156	25	smtp	CVE-2023-3087 - The FluentSMTP plugin for WordPress is vulnerable to Stored Cross-Site Scripting via an email subject in versions up to, and including, 2.2.4 due to insufficient input sanitization and output escaping. This makes it possible for unauthenticated attackers to inject arbitrary web scripts in pages that will execute whenever a user accesses an injected page.	MEDIUM	6.1	CVE-2023-3087, CVE-2024-9065	Schedule remediation within 30 days. Apply vendor patches and security updates. Keep this service patched and secured.
45.33.32.156	80	http Apache httpd 2.4.7 ((Ubuntu))	PORT-80 - Open http Apache httpd 2.4.7 ((Ubuntu)) port detected.	CRITICAL	9.5	CVE-2021-44228, CVE-2022-22965	Schedule remediation within 30 days. Apply vendor patches and security updates. Secure HTTP service with proper authentication, apply web server patches, consider WAF.
45.33.32.156	445	microsoft-ds	PORT-445 - Open microsoft-ds port detected.	HIGH	8.0	CVE-2017-0144, CVE-2020-0796	Immediate mitigation required. Apply security patches, implement firewall rules, or disable service if not required. Disable if not needed, implement strict firewall rules, keep patched.

3. Summary

We successfully delivered a full pipeline that—from a user-entered IP—performs live network scanning, automatically identifies likely vulnerabilities, ranks them by severity, and provides AI-generated remediation steps. Key accomplishments:

- Automated CVE data engineering and AI model training
- Real-time integration of Nmap scanning with NLP matching
- A professional browser-based UI with progress feedback
- Modular, well-documented code for future extension (e.g. live threat feeds integration)

Future work could include: risk scoring by asset importance, support for UDP scans, and a chatbot interface for interactive remediation guidance.

---

4. References

[1] National Vulnerability Database (NVD), "CVE JSON Feeds," https://nvd.nist.gov/feeds/json/cve/1.1/.
[2] "python-nmap: Nmap port scanner wrapper for Python," PyPI, https://pypi.org/project/python-nmap/ Hugging Face.
[3] J. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," EMNLP 2019.
[4] "all-MiniLM-L6-v2," Hugging Face, https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2 Hugging Face.
[5] "T5-small," Hugging Face, https://huggingface.co/google-t5/t5-small Nmap.
[6] Nmap Reference Guide, "Nmap Network Scanning," nmap.org/book/man.html