Project Report

Vulnerability Assessment & Reverse Engineering

Automated Vulnerability Scanner with AI-Powered Classification and Remediation

Team Members: Muhammad Umer Farooq, Jalil Ahmad, Aleena Fatima

Roll Numbers: I221661, I221635, I222353

Class: CY-D

Contents

1. Introduction

1.1 Background

As cyber threats grow in complexity and frequency, automated systems that not only detect but also assess and remediate vulnerabilities are critical. Traditional vulnerability scanners like OpenVAS (Greenbone Vulnerability Management) are effective at detection but lack automated reasoning and response generation capabilities.

1.2 Objective

This project aims to build a comprehensive vulnerability scanning framework that integrates:

- OpenVAS for detection
- Machine learning for severity classification
- Natural language generation (NLG) for remediation recommendations

The system streamlines the vulnerability management process, providing actionable insights in addition to detection.

2. System Overview

2.1 Core Modules

The system is composed of three core modules:

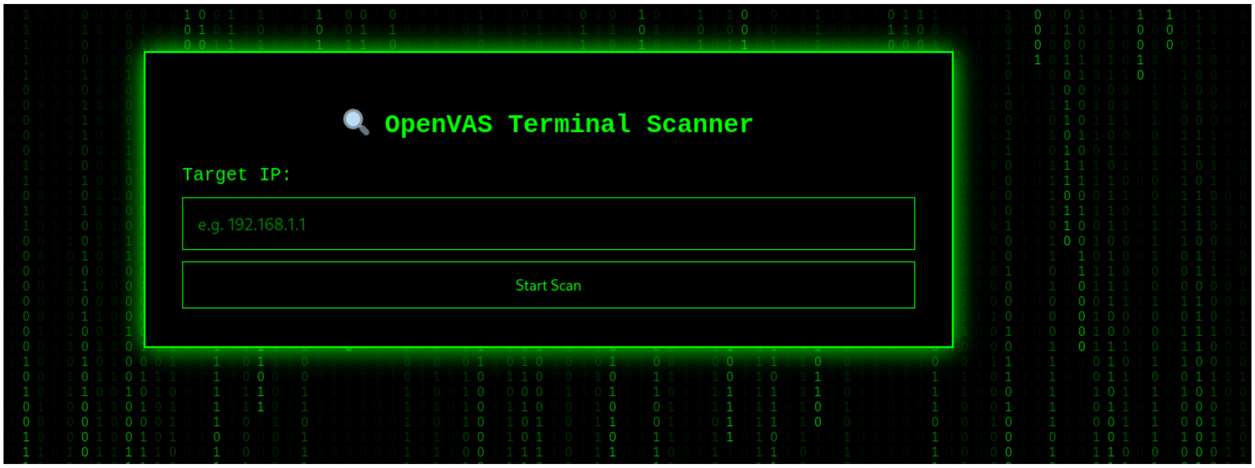
Module	**Function**
Vulnerability Scanning	Conducts system scans using OpenVAS
Severity Classification	Predicts vulnerability severity using ML models
Remediation Recommendation	Generates remediation steps via a fine-tuned T5 model

2.2 Pipeline

Target IP → [OpenVAS Scan] → [Cleaned Report] → [Severity Classifier] → [Remediation Generator] → [Final Output]

3. Technical Components

3.1 Vulnerability Scanning



3.1.1 Tool Used: OpenVAS

OpenVAS is an open-source vulnerability scanner that performs a wide range of network checks against known CVEs.

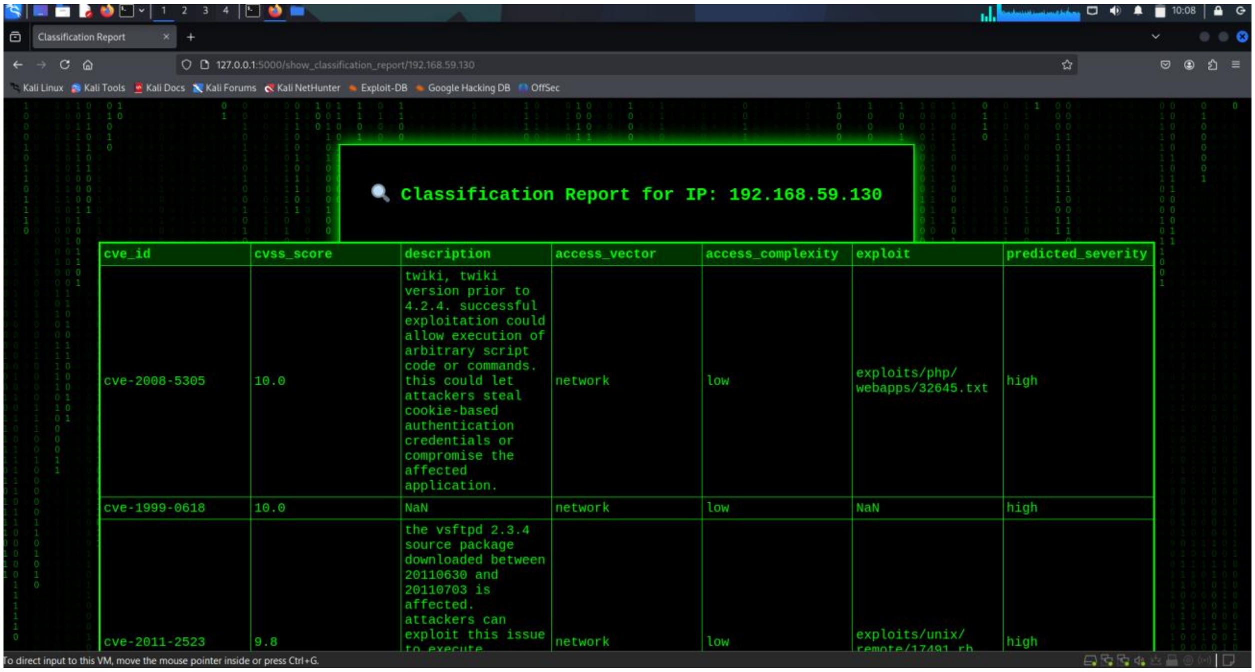
3.1.2 Workflow

- Scan tasks are initiated using gvm-cli via Unix socket.
- The scan is configured to use "Full and fast" policy.
- Results are fetched in CSV format and parsed using pandas.

3.1.3 Data Enrichment

- Extracted data includes CVE ID, CVSS score, solution, impact, and affected software.
- A secondary lookup (if available) is used to enrich the data with:
  - access_vector
  - access_complexity
  - exploit

3.2 Severity Classification



3.2.1 Problem Framing

Severity classification is treated as a supervised classification task with multi-class output (e.g., Low, Medium, High).

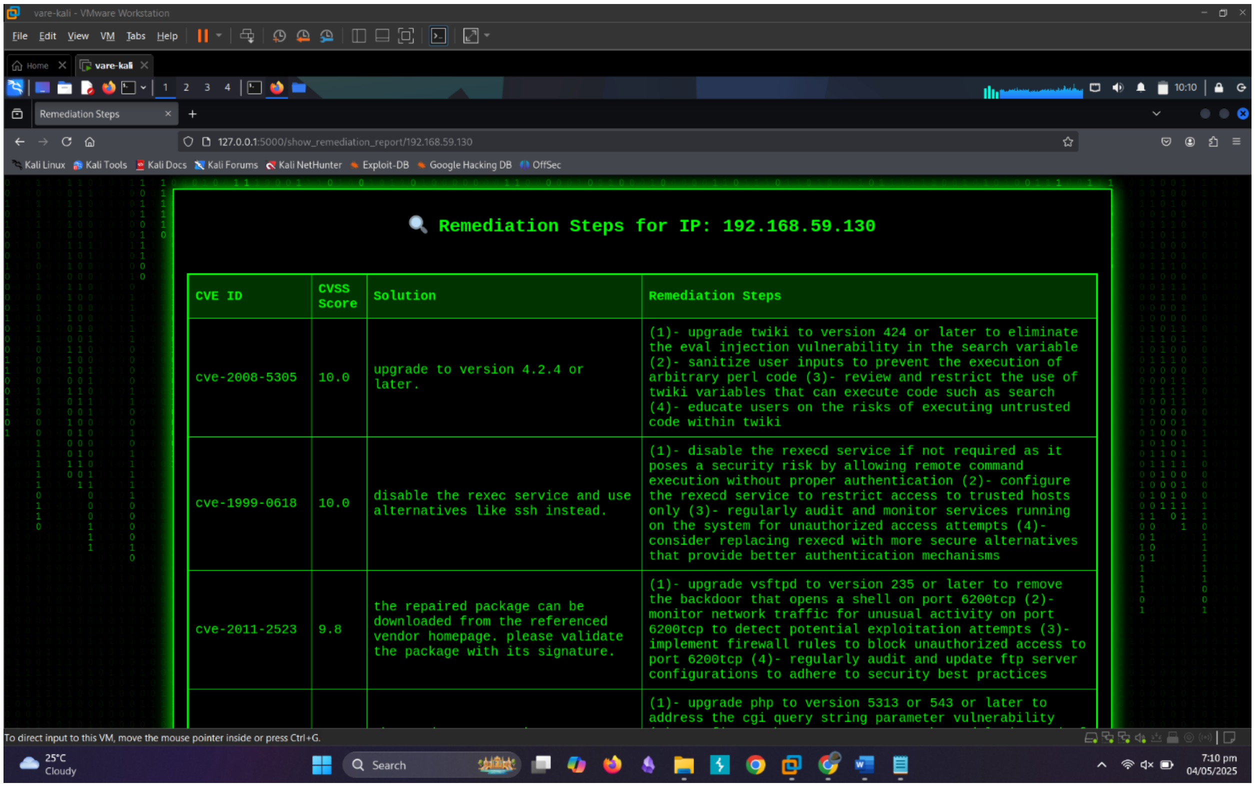
3.2.2 Model Architecture

- **Base Classifier**: Random Forest
- **Text Embedding**: SBERT (all-MiniLM-L6-v2) is used to convert descriptions to numerical vectors.

3.2.3 Feature Set

- Encoded categorical fields: access_vector, access_complexity, exploit
- Continuous fields: cvss_score
- Text embeddings: SBERT vectors from the vulnerability description

3.3 Remediation Generation



Remediation is not fetched from a database but generated using a **fine-tuned T5 (Text-to-Text Transfer Transformer)** model. The model generates contextual, human-readable remediation steps tailored to each CVE and its description.

4. Implementation Details

4.1 Backend

- **Framework**: Flask
- **Task Management**: Dictionary-based tracking of running/completed scans
- **Data Handling**: pandas, joblib, torch, transformers

4.2 Frontend

Implemented using HTML/CSS/JavaScript with a consistent hacker-themed visual identity (Matrix-style green-on-black). Each stage has a dedicated UI:

- index.html: Scan initiation
- results.html: Scan data

- classification.html: Severity predictions
- remediations.html: Final output with recommendations

5. User Workflow

Step Action

1 User enters target IP on the web interface

2 OpenVAS scan begins in background

3 Upon completion, results are shown in a table

4 User can trigger severity classification

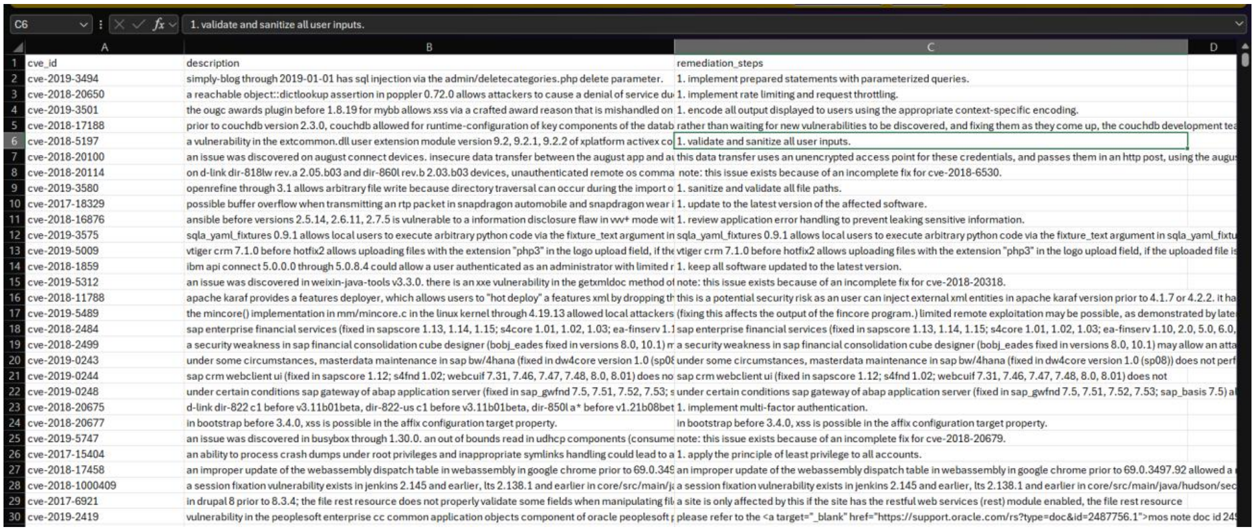
5 Final step generates AI-powered remediation suggestions

6. Datasets Utilised

The data for training severity classification model is gathered from NVD database and ExploitDB for corresponding exploits. A corpus of 200,000 entries is created. Following is the glimpse of it:

cve_id	description	access_vector	access_complexity	exploit	cvss_score	severity
cve-1999-0095	the debug command in sendma	network	low	exploits/lin	10	critical
cve-1999-0082	cwd ~root command in ftpd allo	network	low	null	10	critical
cve-1999-1471	buffer overflow in passwd in bsc	local	low	null	7.2	high
cve-1999-1122	vulnerability in restore in sunos	local	low	null	4.6	medium
cve-1999-1467	vulnerability in rcp on sunos 4.0	network	low	null	10	critical
cve-1999-1506	vulnerability in smi sendmail 4.(	network	low	null	7.5	high
cve-1999-0084	certain nfs servers allow users t	local	low	null	7.2	high
cve-2000-0388	buffer overflow in freebsd libmy	network	low	null	7.5	high
cve-1999-0209	the sunview (suntools) selectior	network	low	exploits/sc	5	medium
cve-1999-1198	builddisk program on next syste	local	low	null	7.2	high
cve-1999-1391	vulnerability in next 1.0a and 1.(	local	low	null	7.2	high
cve-1999-1392	vulnerability in restore0.9 instal	local	low	null	7.2	high
cve-1999-1057	vms 4.0 through 5.3 allows loca	local	low	null	4.6	medium
cve-1999-1554	/usr/sbin/mail on sgi irix 3.3 anc	local	low	null	2.1	low
cve-1999-1197	tioccons in sunos 4.1.1 does no	local	low	null	7.2	high
cve-1999-1115	vulnerability in the /etc/suid_exe	local	low	null	7.2	high
cve-1999-1258	rpc.pwdauthd in sunos 4.1.1 an	network	low	null	5	medium
cve-1999-1438	vulnerability in /bin/mail in suno	local	low	null	7.2	high
cve-1999-1211	vulnerability in in.telnetd in sunc	local	low	null	7.2	high
cve-1999-1212	vulnerability in in.rlogind in sunc	local	low	null	7.2	high
cve-1999-1194	chroot in digital ultrix 4.1 and 4.(	local	low	exploits/ai:	7.2	high
cve-1999-1193	the "me" user in next nextstep 2.	network	low	null	10	critical
cve-1999-1123	the installation of sun source (sı	local	low	exploits/sc	7.2	high
cve-1999-1034	vulnerability in login in at&t syste	local	low	null	7.2	high
cve-1999-1415	vulnerability in /usr/bin/mail in c	local	low	null	4.6	medium
cve-1999-1090	the default configuration of ncs:	network	low	null	7.5	high

The dataset for remediations is gathered through web scraping and contains remediation steps for over 14000 vulnerabilities. It is as follows:



C6 1. validate and sanitize all user inputs.

	A	B	C	D
1	cve_id	description	remediation_steps	
2	cve-2019-3494	simply-blog through 2019-01-01 has sql injection via the admin/deletecategories.php delete parameter.	1. implement prepared statements with parameterized queries.	
3	cve-2018-20650	a reachable object::dictlookup assertion in poppler 0.72.0 allows attackers to cause a denial of service du	1. implement rate limiting and request throttling.	
4	cve-2019-3501	the ougc awards plugin before 1.8.19 for mybb allows xss via a crafted award reason that is mishandled on	1. encode all output displayed to users using the appropriate context-specific encoding.	
5	cve-2018-17188	prior to couchdb version 2.3.0, couchdb allowed for runtime-configuration of key components of the datab	rather than waiting for new vulnerabilities to be discovered, and fixing them as they come up, the couchdb development tea	
6	cve-2018-5197	a vulnerability in the extcommon.dll user extension module version 9.2, 9.2.1, 9.2.2 of xplatform activex co	1. validate and sanitize all user inputs.	
7	cve-2018-20100	an issue was discovered on august connect devices. insecure data transfer between the august app and au	this data transfer uses an unencrypted access point for these credentials, and passes them in an http post, using the augu	
8	cve-2018-20114	on d-link dir-818lw rev.a 2.05.b03 and dir-860l rev.b 2.03.b03 devices, unauthenticated remote os comma	note: this issue exists because of an incomplete fix for cve-2018-6530.	
9	cve-2019-3580	openrefine through 3.1 allows arbitrary file write because directory traversal can occur during the import o	1. sanitize and validate all file paths.	
10	cve-2017-18329	possible buffer overflow when transmitting an rtp packet in snapdragon automobile and snapdragon wear i	1. update to the latest version of the affected software.	
11	cve-2018-16876	ansible before versions 2.5.14, 2.6.11, 2.7.5 is vulnerable to a information disclosure flaw in vvv+ mode wit	1. review application error handling to prevent leaking sensitive information.	
12	cve-2019-3575	sqla_yaml_fixtures 0.9.1 allows local users to execute arbitrary python code via the fixture_text argument in	sqla_yaml_fixtures 0.9.1 allows local users to execute arbitrary python code via the fixture_text argument in sqla_yaml_fixtu	
13	cve-2019-5009	vtiger crm 7.1.0 before hotfix2 allows uploading files with the extension "php3" in the logo upload field, if the	vtiger crm 7.1.0 before hotfix2 allows uploading files with the extension "php3" in the logo upload field, if the uploaded file is	
14	cve-2018-1859	ibm api connect 5.0.0.0 through 5.0.8.4 could allow a user authenticated as an administrator with limited r	1. keep all software updated to the latest version.	
15	cve-2019-5312	an issue was discovered in weixin-java-tools v3.3.0. there is an xxe vulnerability in the getxmldoc method of	note: this issue exists because of an incomplete fix for cve-2018-20318.	
16	cve-2018-11788	apache karaf provides a features deployer, which allows users to "hot deploy" a features xml by dropping th	this is a potential security risk as an user can inject external xml entities in apache karaf version prior to 4.1.7 or 4.2.2. it ha	
17	cve-2019-5489	the mincore() implementation in mm/mincore.c in the linux kernel through 4.19.13 allowed local attackers	(fixing this affects the output of the fincore program.) limited remote exploitation may be possible, as demonstrated by late	
18	cve-2018-2484	sap enterprise financial services (fixed in sapscore 1.13, 1.14, 1.15; s4core 1.01, 1.02, 1.03; ea-finserv 1.1	sap enterprise financial services (fixed in sapscore 1.13, 1.14, 1.15; s4core 1.01, 1.02, 1.03; ea-finserv 1.10, 2.0, 5.0, 6.0,	
19	cve-2018-2499	a security weakness in sap financial consolidation cube designer (bobj_eades fixed in versions 8.0, 10.1) m	a security weakness in sap financial consolidation cube designer (bobj_eades fixed in versions 8.0, 10.1) may allow an atta	
20	cve-2019-0243	under some circumstances, masterdata maintenance in sap bw/4hana (fixed in dw4core version 1.0 (sp08	under some circumstances, masterdata maintenance in sap bw/4hana (fixed in dw4core version 1.0 (sp08)) does not perf	
21	cve-2019-0244	sap crm webclient ui (fixed in sapscore 1.12; s4fnd 1.02; webcuif 7.31, 7.46, 7.47, 7.48, 8.0, 8.01) does no	sap crm webclient ui (fixed in sapscore 1.12; s4fnd 1.02; webcuif 7.31, 7.46, 7.47, 7.48, 8.0, 8.01) does not	
22	cve-2019-0248	under certain conditions sap gateway of abap application server (fixed in sap_gwfnd 7.5, 7.51, 7.52, 7.53; s	under certain conditions sap gateway of abap application server (fixed in sap_gwfnd 7.5, 7.51, 7.52, 7.53; sap_basis 7.5) al	
23	cve-2018-20675	d-link dir-822 c1 before v3.11b01beta, dir-822-us c1 before v3.11b01beta, dir-850l a* before v1.21b08bet	1. implement multi-factor authentication.	
24	cve-2018-20677	in bootstrap before 3.4.0, xss is possible in the affix configuration target property.	in bootstrap before 3.4.0, xss is possible in the affix configuration target property.	
25	cve-2019-5747	an issue was discovered in busybox through 1.30.0. an out of bounds read in udhcp components (consume	note: this issue exists because of an incomplete fix for cve-2018-20679.	
26	cve-2017-15404	an ability to process crash dumps under root privileges and inappropriate symlinks handling could lead to a	1. apply the principle of least privilege to all accounts.	
27	cve-2018-17458	an improper update of the webassembly dispatch table in webassembly in google chrome prior to 69.0.349	an improper update of the webassembly dispatch table in webassembly in google chrome prior to 69.0.3497.92 allowed a	
28	cve-2018-1000409	a session fixation vulnerability exists in jenkins 2.145 and earlier, lts 2.138.1 and earlier in core/src/main/ja	a session fixation vulnerability exists in jenkins 2.145 and earlier, lts 2.138.1 and earlier in core/src/main/java/hudson/sec	
29	cve-2017-6921	in drupal 8 prior to 8.3.4; the file rest resource does not properly validate some fields when manipulating fil	a site is only affected by this if the site has the restful web services (rest) module enabled, the file rest resource	
30	cve-2019-2419	vulnerability in the peoplesoft enterprise cc common application objects component of oracle peoplesoft p	please refer to the <a target="_blank" href="https://support.oracle.com/rs?type=doc&id=2487756.1">mos note doc id 24	

7. Model Evaluation and Testing

This section presents the evaluation results for both core models developed as part of the vulnerability scanning framework: the **Severity Classification Model** and the **Remediation Generation Model**. Each model was rigorously tested to ensure accuracy, generalization, and utility in a real-world cybersecurity context.

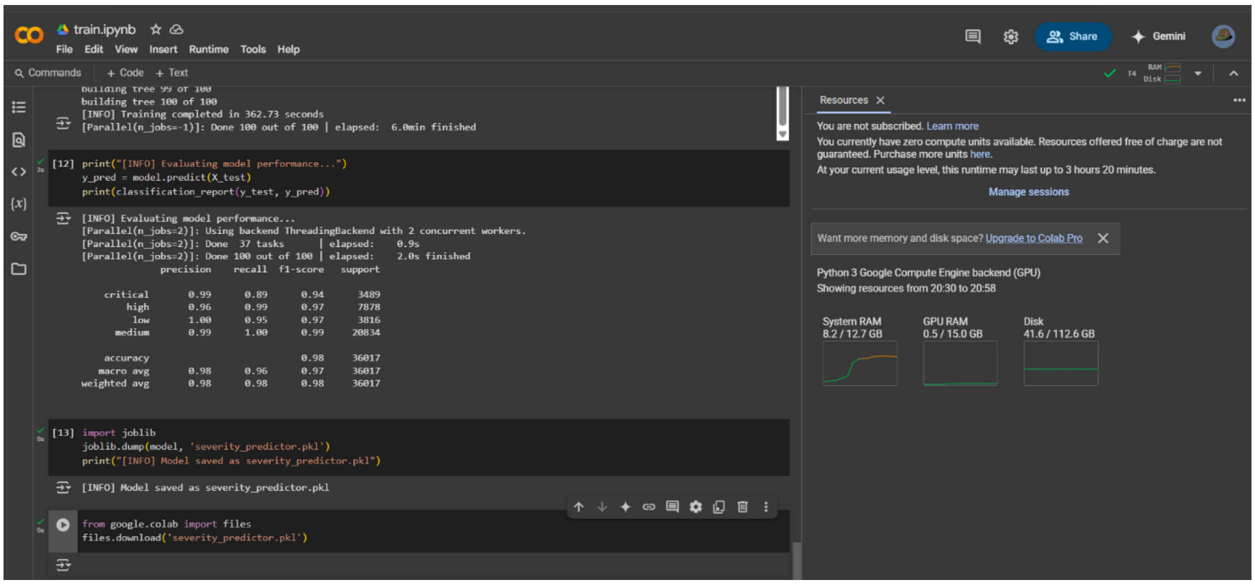
7.1 Severity Classification Model

The severity classification model is a Random Forest classifier trained on enriched CVE metadata. Input features included the cvss_score, access_vector, access_complexity, and SBERT-encoded descriptions of vulnerabilities. The model was trained using 100 trees and evaluated on a held-out test set.

Evaluation Metrics

Class	Precision	Recall	F1-Score	Support
Critical	0.99	0.89	0.94	3489
High	0.96	0.99	0.97	7878
Low	1.00	0.95	0.97	3816
Medium	0.99	1.00	0.99	20834
Accuracy			**0.98**	36017

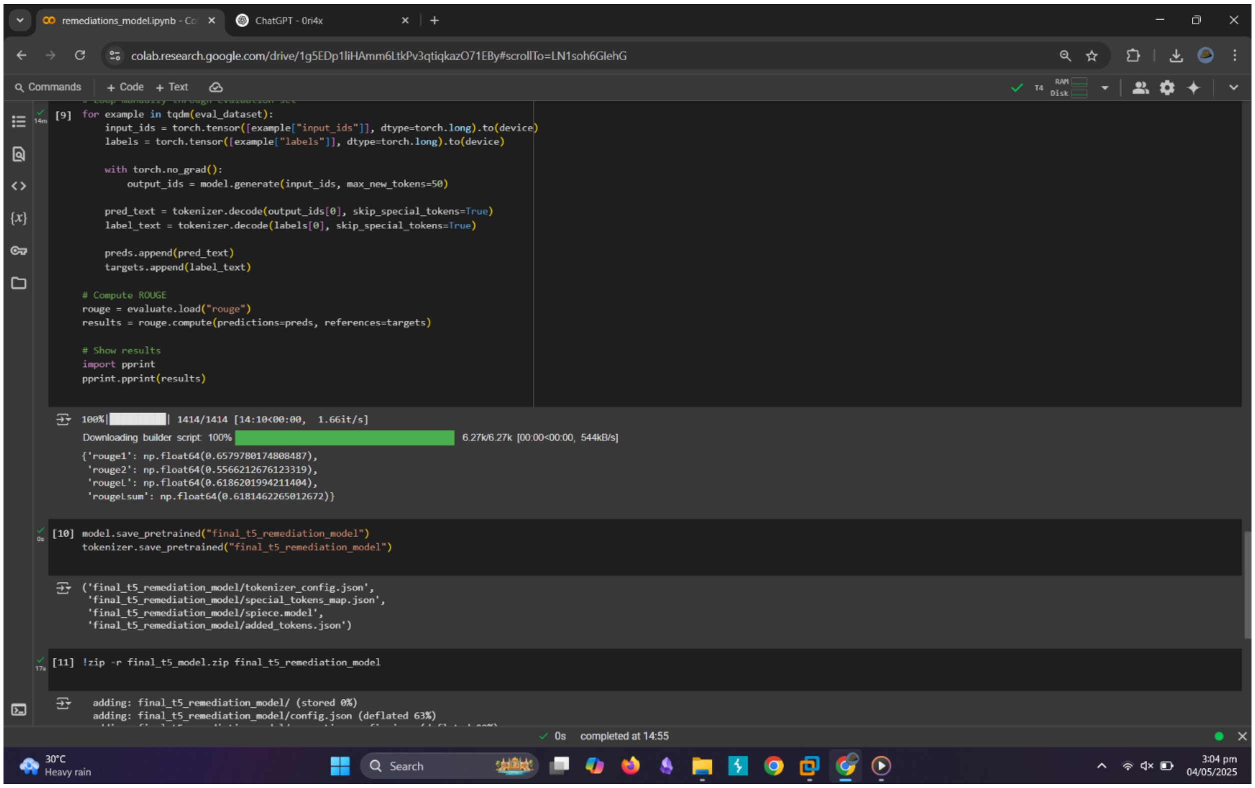
Class	Precision	Recall	F1-Score	Support
Macro Avg	0.98	0.96	0.97	36017
Weighted Avg	0.98	0.98	0.98	36017



7.2 Remediation Generation Model

The remediation model is a fine-tuned T5 transformer trained on a custom dataset containing CVE IDs, descriptions, and corresponding remediation steps. The model is tasked with generating human-readable, context-specific mitigation recommendations from structured inputs.

Metric	Score
ROUGE-1	0.597
ROUGE-2	0.566
ROUGE-L	0.618
ROUGE-Lsum	0.618



8. Limitations and Future Work

8.1 Current Limitations

- OpenVAS may miss zero-day vulnerabilities.
- The classifier's accuracy is bounded by the training data and SBERT embeddings.
- The T5 model may generate plausible but non-authoritative remediations.

8.2 Proposed Enhancements

- Integrate NVD API for real-time vulnerability enrichment.
- Replace SBERT with a domain-specific LLM fine-tuned on MITRE CVE data.
- Build a dashboard for multi-host scanning and vulnerability trend analytics.

9. Conclusion

This project bridges the gap between detection and response by combining the strengths of traditional vulnerability scanners and modern AI models. It provides:

- End-to-end automation
- Human-like remediation suggestions
- An intuitive UI to manage scan workflows

This system is a step toward building **smart cybersecurity assistants** that can reason and recommend—automatically.