



جامعة  
Teknologi  
MARA

CSP760

*PREDICTING VULNERABILITY SUSCEPTIBILITY IN  
MALAYSIAN BANK USING SUPERVISED MACHINE  
LEARNING*

STUDENT

SUPERVISOR

NOR ADANI BINTI KAMAL MOHAMAD NASIR (2024782087)

DR SITI ARPAH BINTI AHMAD

Date: 19 October 2025



## TABLE OF CONTENT

- 1 UPDATE WEEK 2
- 2 CHAPTER 1

## UPDATE WEEK 2

Item	Input	Process	Output / Expected Outcome	Status
1. Orange Data Mining Software	Orange Data Mining (Open-source ML tool)	<ul style="list-style-type: none"><li>- Installed and explored Orange Data Mining interface.</li><li>- Familiarized with visual programming workflow (widgets, data input, preprocessing, model training).</li><li>- Tested sample datasets to understand data flow.</li></ul>	<ul style="list-style-type: none"><li>- Software successfully installed and operational.</li><li>- Clear understanding of Orange Data Mining functions for ML model training phase.</li></ul>	<span>✓ Done</span>
2. GitHub Setup	GitHub platform	<ul style="list-style-type: none"><li>- Created GitHub account dedicated to FYP project.</li><li>- Set up initial repository for version control and code storage.</li><li>- Configured structure for dataset, script, and documentation folders.</li><li>- Planned future integration with Streamlit for deployment.</li></ul>	<ul style="list-style-type: none"><li>- GitHub repository ready for collaboration and version tracking.</li><li>- Organized code management system for upcoming development stages.</li></ul>	<span>✓ Done</span>
3. Dataset Findings	<ul style="list-style-type: none"><li>- Open-source datasets (Kaggle, Tenable, CVE, ExploitDB)</li><li>- Internal VA reports (in progress)</li></ul>	<ul style="list-style-type: none"><li>- Collected CVE dataset (✓).</li><li>- ExploitDB dataset collection ongoing.</li><li>- Internal / historical data extraction in progress.</li></ul>	<ul style="list-style-type: none"><li>- CVE dataset ready for use.</li><li>- Comprehensive dataset to be finalized after integrating all sources.</li></ul>	<span>⚙️ In Progress</span>
4. Data Pre-processing & Cleaning	Raw dataset (Kaggle, Tenable, CVE, ExploitDB)	<ul style="list-style-type: none"><li>- Started cleaning raw data (handling missing values, duplicates, data type normalization).</li><li>- Began preliminary Exploratory Data Analysis (EDA) using Orange.</li><li>- Feature identification for model training (Plugin ID, Severity, CVSS, CVE, Family).</li></ul>	<ul style="list-style-type: none"><li>- Clean dataset under development.</li><li>- Early data insights and visualization obtained.</li></ul>	<span>⚙️ In Progress</span>
5. Report Update (Chapter 3 Revision)	Supervisor's feedback	<ul style="list-style-type: none"><li>- Revised Chapter 3 based on comments.</li><li>- Moved theoretical explanation to Chapter 2.</li><li>- Reorganized methodology to emphasize step-by-step process and tools.</li></ul>	<ul style="list-style-type: none"><li>- Updated report aligned with supervisor's feedback.</li><li>- Improved chapter structure and clarity of methodology section.</li></ul>	<span>⚙️ In Progress</span>

# INTRODUCTION

Cyber threats are rising rapidly in Malaysia's banking sector. Banks rely on Vulnerability Assessment (VA) tools, which only detect existing issues and not the future ones. This reactive approach leaves banks unprepared for new attack patterns. To address this, Machine Learning (ML) can analyze past VA data to predict potential vulnerabilities and improve threat mitigation. This study aims to use ML to make cybersecurity in Malaysian banks more proactive and predictive.

## RESEARCH QUESTIONS

- What are the limitations of current VA tools in predicting future attacks?
- Can ML models improve how banks prioritize and fix vulnerabilities by forecasting attack trends?
- How effective is the ML-based model compared to traditional VA methods?

## HYPOTHESIS

This study proposes that machine learning can enhance the predictive capability of vulnerability assessment in Malaysian banks.

## RESEARCH OBJECTIVES

- Design an ML model to predict cyberattack susceptibility using VA data.
- Develop and test the ML system for predicting future vulnerabilities.
- Evaluate the model's accuracy and effectiveness in forecasting potential attack

# RESEARCH CHAPTER 1

## PROBLEM STATEMENT

- Current vulnerability assessment (VA) tools in Malaysian banks identify existing vulnerabilities but fail to provide predictive insights into future attack trends (Hanif et al., 2021), leaving organizations reactive rather than proactive in managing security risks (Gencer & Başçiftçi, 2021).
- Manual analysis of large-scale VA results is time-consuming, error-prone, and inefficient (Esposito & Falessi, 2024), hindering the ability to prioritize critical vulnerabilities effectively (He et al., 2024).
- The rapidly evolving nature of cyber threats needs an intelligent system such as, AI-driven system to predict potential attack vectors and mitigate risks before exploitation occurs (Zacharis et al., 2024).

## SIGNIFICANCE OF STUDY

1. Knowledge Contribution
  - Fills research gaps in predictive cybersecurity for Malaysian banks.
  - Shifts from reactive to proactive threat management.
2. AI-Based Predictive Model
  - Reduces manual effort and time.
  - Enhances vulnerability prioritization efficiency.
3. Industry Impact
  - Strengthens financial institutions' cybersecurity posture.
  - Promotes adoption of Industry 4.0 and AI innovation.
4. Proactive Risk Management
  - Enables early identification of potential attack vectors.
  - Supports continuous improvement in risk remediation.

## RESEARCH SCOPE

1. Uses Supervised ML models:
  - a. Random Forest (Classification)
  - b. Neural Network Regression (Prediction)
2. Data sources: Kaggle and Tenable datasets.
3. Key attributes: Plugin ID, Severity, CVSS, Findings, CVE, Asset Type, Family.



UNIVERSITI  
TEKNOLOGI  
MARA

CSP760

THANK YOU