National College of Ireland

# AI-Driven Automated Cyber Risk Assessment Pipeline for GRC Intelligence – A Tripartite Framework

MSc Research Project
Data Analytics

Kaviya Maniyan
x23377879

School of Computing
National College of Ireland

Supervisor: Prof. Arjun Chikkankod

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** Kaviya Maniyan

**Student ID:** X23377879@student.ncirl.ie

**Programme:** Data Analytics **Year:** 2025

**Module:** MSC Research Project

**Supervisor:** Prof. Arjun Chikkankod

**Submission Due Date:** 11/12/2025

**Project Title:** AI-Driven Automated Cyber Risk Assessment Pipeline for GRC Intelligence – A Tripartite Framework

**Word Count:** 473 **Page Count:** 3

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:** Kaviya Maniyan

**Date:** 10/11/2025

**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

Kaviya Maniyan
x23377879

**Title:** AI-Driven Automated Cyber Risk Assessment Pipeline for GRC Intelligence – A Tripartite Framework

# 1 Introduction

The configuration manual provides a step-by-step process to install, configure and operate the developed cybersecurity system of this research. This manual provides clear instructions to run, reproduce results and required systems for setting up the environment. The system integrates both structured and unstructured CVE vulnerability datasets, semantic mapping embeddings, trained models and knowledge graph. Users following this manual can execute an automated interactive assistant that predicts the severity of vulnerability then maps with related attack techniques and standard policy controls.

# 2 System Requirements

## Hardware

- Minimum 8 GB Ram
- For faster model inference/computation and quick embedding use NVIDIA GPU or Google Colab Pro
- Free storage space of approximately 10 – 15 GB

## Operational System

Windows, macOS or Linus

## Software stack

- Python 3.10 or above
- Jupyter Notebook or Google Colab Pro

# 3 Required Libraries

- pip install sentence-transformers
- pip install scikit-learn
- pip install lightgbm xgboost
- pip install pandas numpy
- pip install tensorflow

- pip install networkx
- pip install joblib
- pip install matplotlib seaborn
- pip install imbalance-learn

**Note**: Some environments use/install imbalanced-learn as imblearn.

# 4 Dataset requirements

Place the datasets below in working directory (or update the paths accordingly):

All the datasets must include text descriptions and identifiers

| | |
|---|---|
| National Vulnerability Database (NVD)- CVE dataset with structured CVSS metrics + text descriptions | nvd_data_COPY.csv |
| MITRE ATT&CK dataset (enterprise attack JSON converted to CSV file) | MITRE_ATT&CK_Enterprise_Techniques.csv |
| NIST SP 800-53 Rev.5 dataset (OSCAL JSON converted to CSV file) | NIST_SP800-53_Rev5_with_subcontrols.csv |
| Semantic mapping datasets | CVE_to_MITRE_top3.csv,<br>MITRE_to_NIST_Top3.csv,<br>CVE_to_NIST_Top3.csv |

# 5 Embedding Files

The system uses automatically generated Sentence-BERT embedding files which reduces computation time for each run:

- embeddings_cve.npy
- embeddings_cve.pt
- embeddings_mitre.npy
- embeddings_nist.npy
- mitre_embeddings.pt
- nist_full_embeddings.pt

# 6 Model Configuration

The hybrid severity prediction model integrates structured CVSS features and text descriptions → MLP (deep neural classifier), the pipeline must exist:

- mlp_sentencebert_fusion.keras
- preprocessor. Joblib

# 7 Running the system

- Open the nvd_colab.ipynb
- Ensure the working directory, file paths of datasets, run the hybrid MLP model, embeddings, mapping CSVs
- Then execute the assistant cell which shows a text prompt, provide either CVE ID as a string or a free text.

**Example inputs:** CVE-2020-0001 or SQL injection vulnerability in login form allowing unauthorized data access.

# 8 Output Interpretation

The system produces:

- Predicted severity and the probability distribution of all cases (LOW, MEDIUM, HIGH and CRITICAL)
- Top 3 MITRE ATT&CK techniques and Top 3 NIST controls with cosine similarity score
- Also top 3 NIST controls for each MITRE technique
- A tri-partite subgraph visualization

# 9 Troubleshooting

Missing embeddings → Run the embedding cell to regenerate automatically

Slow execution → Enable GPU in Google Colab Pro (Runtime → Change runtime → GPU)

Mapping not found → Verify file name and file paths

Missing model or preprocessor → Re-run the training notebook