Project Title: APT DETECTION AI

Attacks Handled: DDOS Attacks

Dataset: Network Intrusion dataset(CIC-IDS- 2017)

Prepared By: [Khadijah Farooqi]

Date: [2nd May 2025]

---

1. Objectives

Ensure that the machine learning pipeline:

- Delivers high detection accuracy.
- Handles noisy or incomplete data gracefully.
- Is robust, maintainable, and scalable.

---

2. Standards to Follow

- **PEP8** for Python code formatting.
- **CRISP-DM** methodology for ML lifecycle.
- **IEEE 829** for test documentation (adapted).
- **Data privacy** ensured by anonymizing input samples.

---

3. SQA Activities

Activity	Description	Tools
Code Review	Manual and/or automated check for code consistency, variable naming, imports, etc.	VSCode, pylint
Data Validation	Ensure datasets contain required features, correct formats, and manageable missing values.	pandas, numpy

Activity	Description	Tools
Unit Testing	Validate each function/module (e.g., transformation, prediction)	pytest
Functional Testing	End-to-end test using test.csv and checking prediction accuracy	Python scripts
Performance Testing	Measure model latency and memory usage on large files	time, memory-profiler
Regression Testing	Re-check accuracy on known samples after changes	`train_val.csv`

---

4. Test Plan

Test Case	Description	Expected Result
TC01	Load model from disk	Model loaded without errors
TC02	Validate feature presence	Missing columns handled correctly
TC03	Apply transformation to incomplete data	Output has full 59 features
TC04	Predict using clean test set	Accuracy > 99%
TC05	Predict using synthetic set	Accuracy drop is expected but code runs
TC06	Check predictions saved to file	`predictions_with_labels.csv` generated

---

5. Tools & Environment

- Python 3.11
- scikit-learn
- pandas, numpy
- joblib
- OS: Parrot Security OS
- VSCode

---

6. Responsibilities

- **Data Preprocessing** – [Khadijah]
- **Model Development** – [Khadijah]
- **QA/Testing** – [Khadijah]