



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

INDEPENDENT PROJECT ***(ACCA6020)***

Muhammad Kamrul Hasan Sohel

MODULE CODE: ACCA6020

STUDENT ID: 2316411

LECTURER NAME: RENUGA JAYAKUMAR

UNIVERSITY OF WALES TRINITY SAINT DAVID

Table of Contents

1. Introduction / Background to the Proposed Research.....	1
Research Aim and Objectives	2
2. Project Design.....	3
3. Resources and Planning	4
Conclusion	5
References.....	6

PROJECT PROPOSAL

AI-Based Threat Detection and Alert System for Cloud Security

1. Introduction / Background to the Proposed Research

Working title: AI-Based Threat Detection and Alert System for Cloud Security. The proposed project investigates how machine learning can support cyber security teams by analysing network traffic records and identifying potentially malicious activity. Cloud services are widely used for storage, applications and remote working, but they also increase exposure to attacks such as denial of service, reconnaissance, exploitation, malware communication and unauthorised access. Traditional rule-based intrusion detection can miss new or unusual patterns because it depends on predefined signatures. This project therefore proposes a simple AI-enabled threat detection system that learns from labelled cyber security data and presents the detected risks through a web dashboard.

The research is relevant because recent threat reporting continues to show that cyber-attacks remain a major operational risk. ENISA identifies denial-of-service, malware, data compromise, social engineering and vulnerability exploitation as continuing threat categories in the European landscape [1]. NIST also frames detection, response and recovery as core functions within cyber security risk management [2]. For a small student prototype, it is not

realistic or ethical to monitor a live cloud environment. Instead, the project will use an established benchmark dataset, UNSW-NB15, which contains normal traffic and nine attack families including Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms [3].

A brief literature review shows that machine learning is commonly used in network intrusion detection because classification algorithms can learn relationships between traffic features and attack labels. The UNSW-NB15 dataset remains popular because it contains modern normal and attack behaviours with extracted network features [3], while the CICIDS2017 dataset was designed to include benign traffic and common real-world attack scenarios in labelled flow format [4]. Research on UNSW-NB15 also highlights practical issues such as class imbalance and class overlap, meaning that a model can appear accurate overall while still performing poorly for minority attack types [5]. This creates a clear research gap for a student project: to develop a transparent, lightweight prototype that not only trains a model but also evaluates performance using suitable metrics and presents the result in a usable dashboard.

Research Aim and Objectives

The aim of the study is to design and develop a simple AI-based cyber security system that can analyse network traffic data, classify activities as normal or malicious, and generate alerts through a web-based dashboard.

Objective	Project Activity	Expected Evidence
O1	Investigate common cyber security threats and intrusion detection methods.	Short literature review and concept summary.
O2	Obtain and analyse UNSW-NB15 or a similar labelled dataset.	Dataset description and data dictionary.
O3	Clean, encode and scale the dataset for machine learning.	Pre-processing notebook and prepared dataset.
O4	Train a Random Forest model to classify normal and malicious records.	Saved model and training results.
O5	Build a Flask dashboard to upload data and show detections.	Working web prototype screenshots.
O6	Evaluate the model using accuracy, precision, recall,	Evaluation table, charts and discussion.

	F1-score and confusion matrix.	
--	--------------------------------	--

2. Project Design

The project will follow an applied development design combined with experimental evaluation. The artefact will be a small web-based prototype rather than a full enterprise security platform. The product development element will cover the dashboard, database and alert interface. The experimental element will test whether a supervised machine learning classifier can distinguish normal and malicious activity using labelled dataset records.

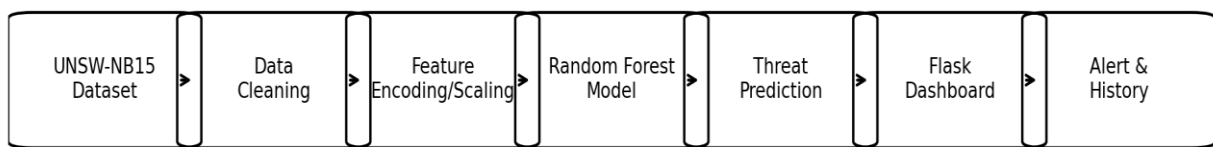


Figure 1. Proposed system workflow.

Data collection will use secondary data only. The main dataset will be UNSW-NB15 because it is openly available for research and provides labelled attack categories. The dataset will be inspected for missing values, duplicates, feature types, class distribution and irrelevant fields. Categorical variables such as protocol, service and state will be encoded. Numerical fields will be scaled where required. The dataset will then be split into training and test sets so that the final performance is measured on unseen records rather than the same records used for training.

The chosen development method is a simple iterative prototype. First, the machine learning notebook will be built and tested. Second, a Flask application will be created with pages for data upload, dashboard summary, prediction results and alert history. Third, an SQLite database will store prediction results so that the dashboard can show previous alerts. A Random Forest classifier is selected because it is suitable for tabular data, can handle non-linear relationships and gives feature importance values that help explain which traffic features contribute to decisions. If time allows, Logistic Regression or Decision Tree may be trained as baseline models for comparison.

The system will be evaluated using accuracy, precision, recall, F1-score and a confusion matrix. Recall is particularly important because missing a malicious record may be more damaging than incorrectly flagging one normal record. The analysis will also discuss false positives, false negatives and whether the results can be generalised beyond the benchmark dataset. Generalisability is limited because benchmark data does not fully represent live cloud systems; however, the prototype will demonstrate a valid research and development process that can later be extended to real NetFlow or cloud log data.

Risk / Issue	Impact	Mitigation	Ethical Position
Dataset imbalance	Weak detection of minority attacks	Use stratified split, confusion matrix and F1-score	Report limitations honestly
False alerts	Dashboard may overstate risk	Label alerts as prototype outputs	No automated real-world blocking
Data privacy	Live data may contain sensitive identifiers	Use public benchmark data only	No personal data collection
Model misuse	Tool could be misunderstood as production security	State scope and limitations	Educational prototype only

Ethical, legal and professional issues are addressed by limiting the project to public research datasets and avoiding collection of personal or live organisational traffic. The dashboard will not make automatic security decisions; it will only display predicted classifications and alert messages. The project will follow responsible AI principles by documenting the dataset, evaluation metrics, limitations and possible bias. NIST AI risk management guidance emphasises that AI risks should be identified, measured and managed across the system lifecycle [6].

3. Resources and Planning

The resources required are realistic for a student project. Hardware will include a standard laptop or university computer. Software will include Python, Jupyter Notebook, pandas, NumPy, scikit-learn, Flask, SQLite, HTML, CSS and Bootstrap. The main outputs will be the research proposal, literature review notes, pre-processing notebook, trained model, Flask dashboard, evaluation charts, screenshots and final project report.

Resource	Purpose	Justification
Python / Jupyter	Data analysis and model training	Widely used for machine learning experiments
Scikit-learn	Random Forest model and metrics	Appropriate for tabular supervised learning
Flask	Web application dashboard	Lightweight and easier than a full framework
SQLite	Store detection history	Simple local database for prototype

UNSW-NB15	Training and testing data	Recognised benchmark IDS dataset
-----------	---------------------------	----------------------------------

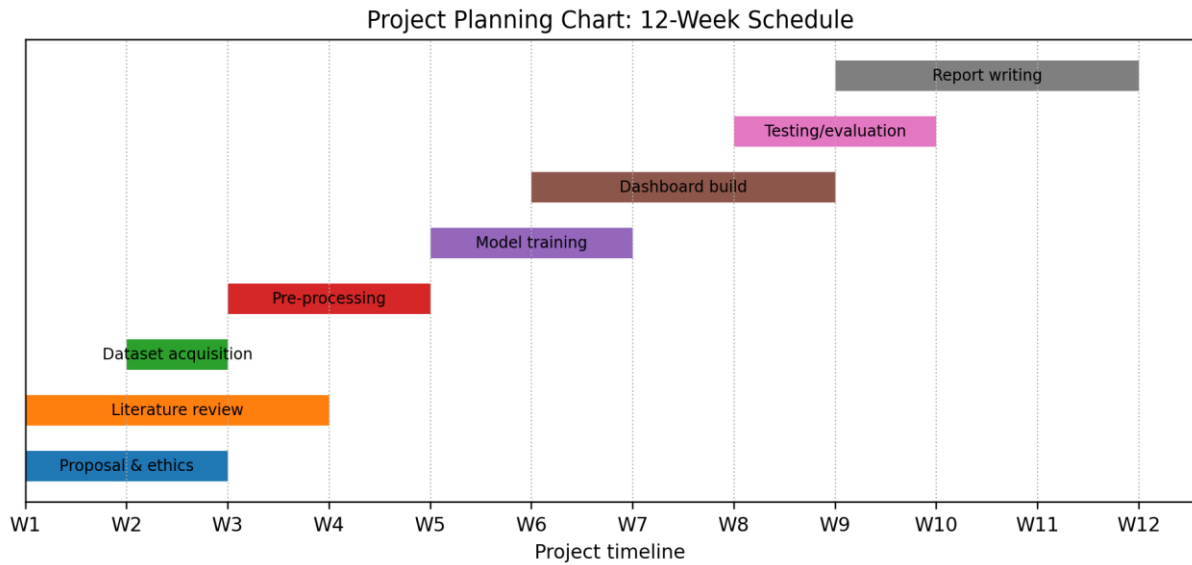


Figure 2. Project planning chart showing timing and resource use.

The systematic literature review process will use IEEE Xplore, ScienceDirect, Google Scholar and university library databases. Search terms will include “AI intrusion detection”, “machine learning network security”, “UNSW-NB15 Random Forest”, “cloud security threat detection” and “cyber security dashboard”. Inclusion criteria will prioritise peer-reviewed papers, recognised datasets and official guidance. PRISMA 2020 will be used as a reporting guide because it supports transparent identification, screening and selection of literature [7]. The review will also consider NetFlow-based intrusion detection datasets for improving future generalisability [8], recent UNSW-NB15 feature-selection and deep-learning research for comparison with the selected Random Forest approach [9], and practical machine-learning guidance for preprocessing, model training and evaluation [10].

Conclusion

This proposal presents a focused and achievable independent research project. It addresses a relevant cyber security problem, uses recognised secondary data, applies appropriate machine learning methods and produces a practical dashboard prototype. The project is limited in scope but strong enough to demonstrate literature-informed research, technical development, ethical awareness and professional evaluation.

References

- [1] ENISA, ENISA Threat Landscape 2024, European Union Agency for Cybersecurity, 2024.
- [2] National Institute of Standards and Technology, The NIST Cybersecurity Framework (CSF) 2.0, NIST, 2024.
- [3] University of New South Wales, “The UNSW-NB15 Dataset,” UNSW Canberra Cyber, 2021.
- [4] Canadian Institute for Cybersecurity, “Intrusion Detection Evaluation Dataset (CICIDS2017),” University of New Brunswick, 2017.
- [5] Z. Zoghi and G. Serpen, “UNSW-NB15 Computer Security Dataset: Analysis through Visualization,” arXiv, 2021.
- [6] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST, 2023.
- [7] M. J. Page et al., “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” BMJ, vol. 372, 2021.
- [8] M. Sarhan, S. Layeghy, N. Moustafa and M. Portmann, “NetFlow Datasets for Machine Learning-Based Network Intrusion Detection Systems,” arXiv, 2020.
- [9] M. Jouhari, H. Benaddi and K. Ibrahimi, “Efficient Intrusion Detection: Combining Feature Selection with CNN-BiLSTM on the UNSW-NB15 Dataset,” arXiv, 2024.
- [10] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow, 3rd ed., O’Reilly Media, 2022.