

Assignment 1: Innovation & Project Management

04 February 2026 10:37

Module: The Changemakers: Creativity & Value Creation

Assignment: Innovation & Project Management

Student Name: Leo Richards

Student Number: 2418519

Course: Cyber Security & networking

Task 1

04 February 2026 10:35

Innovation

Innovation plays a crucial role in shaping modern society by introducing new ideas, technologies, and methods that improve how individuals and organisations operate. Innovation is the process of transforming creative ideas into practical solutions that generate value for society, businesses, and individuals. In the digital age, innovation is strongly linked to technological advancement. [1], particularly in fields such as artificial intelligence, cybersecurity and data analytics.

Creativity is a fundamental component of innovation. Creativity involves generating original ideas, while innovation focuses on implementing those ideas into real-world solutions. Organisations that encourage creativity are often better positioned to develop innovative products, services and processes. For example, technology companies frequently invest in research and development to produce innovative systems that improve security, efficiency and productivity.

Innovation also contributes significantly to economic growth and social development. Many modern technologies, such as cloud computing, machine learning and data analytics, have transformed industries by improving decision-making, automation and service delivery. In cybersecurity, innovation is essential given the constant evolution of cyber threats. Cybercriminals continually develop new attack methods, so security professionals must innovate to protect digital infrastructure. [2]

Phishing attacks are among the most common cyber threats affecting organisations and individuals worldwide. Phishing involves attackers sending fraudulent emails that appear legitimate to trick users into revealing sensitive information, such as passwords, financial details, or personal data. According to cybersecurity reports, phishing attacks account for a significant share of data breaches and financial losses globally.

Innovative cybersecurity solutions are therefore required to detect and prevent phishing attempts more effectively. Machine learning technologies have emerged as a promising approach for identifying phishing emails by analysing patterns in large datasets. By training machine learning algorithms to recognise phishing email characteristics, security systems can automatically classify emails as legitimate or malicious. This type of innovation improves digital security by reducing the reliance on manual detection and improving the speed and accuracy of threat identification. [3]

Overall, innovation is essential to addressing modern technological challenges, particularly in cybersecurity. By combining creativity, research and technological tools, innovative solutions can be developed to protect organisations and individuals from emerging digital threats.

Ideation

Several potential project ideas related to cybersecurity and data analysis were considered.

The first idea was to develop a password-strength analysis tool that evaluates password complexity and provides recommendations to improve password security. This tool would analyse password length, character variation and common password patterns.

The second idea was to create a cybersecurity awareness training platform for small businesses. The platform would provide educational resources, interactive learning modules and simulated phishing exercises to help employees recognise cyber threats.

The third idea was to build a machine-learning-based phishing email detection system that automatically identifies phishing emails by their characteristics. This idea focuses on analysing datasets of legitimate and malicious emails to train a machine learning model that accurately classifies emails.

After evaluating these ideas, the third was selected because phishing attacks remain among the most widespread cybersecurity threats. Additionally, machine learning techniques offer an innovative approach to improving phishing detection.

Project Title

Machine Learning-Based Phishing Email Detection System

Aim

This project aims to develop a machine-learning-based system to detect phishing emails and enhance users' cybersecurity.

Objectives

1. To investigate the characteristics and techniques used in phishing attacks.
2. To collect and analyse datasets containing phishing and legitimate emails.
3. To develop a machine learning model capable of classifying emails as phishing or legitimate.
4. To evaluate the performance of the phishing detection system using testing data.
5. To present the results using data visualisation techniques.

Project Description

This project will focus on developing an intelligent phishing-detection system using machine-learning techniques. The system will analyse email features such as message content, links, sender information and other metadata to identify potential phishing attempts. By training a machine learning model using labelled datasets, the system will learn patterns associated with phishing emails. Once trained, the model will automatically classify incoming emails as phishing or legitimate.

The project aims to demonstrate how machine learning can address cybersecurity challenges and how automated detection systems can enhance digital safety.

Research Sources

Several key sources will support the development of this project.

Academic research papers provide valuable insight into machine learning approaches for phishing detection. Studies published in cybersecurity journals explain how algorithms such as decision trees, support vector machines, and neural networks can detect phishing emails.

Cybersecurity frameworks and reports from organisations such as the National Institute of Standards and Technology (NIST) [4] Also, guide cybersecurity best practices and

threat detection techniques.

Online datasets such as those available on Kaggle will be used to obtain labelled phishing and legitimate email data for training and testing the machine learning model. [5]

Additionally, technical documentation for Python libraries such as Scikit-learn and Pandas will help implement machine learning algorithms and perform data analysis in the project. [6] [7]

These sources will provide the theoretical knowledge, technical guidance and datasets required to achieve the project's aim and objectives.

Task 2 - Project Management

04 February 2026 10:36

Project Scope

For this project, the aim is to develop a machine-learning-based phishing email detection system that automatically identifies and classifies phishing emails.

Phishing attacks remain one of the most common cyber threats targeting both individuals and organisations. These attacks attempt to trick users into revealing confidential information such as login credentials, financial details or personal data. The proposed solution to this problem is to develop a system that uses machine learning algorithms to analyse email content and detect characteristics commonly associated with phishing attacks. [8]

The proposed system will analyse features such as email subject lines, message content, embedded links and sender information. By training the model using labelled datasets of phishing and legitimate emails, the system will learn patterns that distinguish malicious emails from legitimate ones. Once trained, the model will automatically classify incoming emails and flag suspicious messages.

Several technologies and resources will be required to implement this project. The primary programming language used will be Python, due to its strong support for data analysis and machine learning. Python libraries such as Pandas and NumPy will be used for data processing, while Scikit-learn will be used to implement machine learning algorithms. The notebook will be used as the development environment for data analysis and model training.

A dataset containing phishing and legitimate emails will be obtained from publicly available sources such as Kaggle. The data set will be cleaned and prepared before being used to train the machine learning model. Data preprocessing techniques such as feature extraction, text vectorisation, and data normalisation will be applied to improve the model's performance. [5]

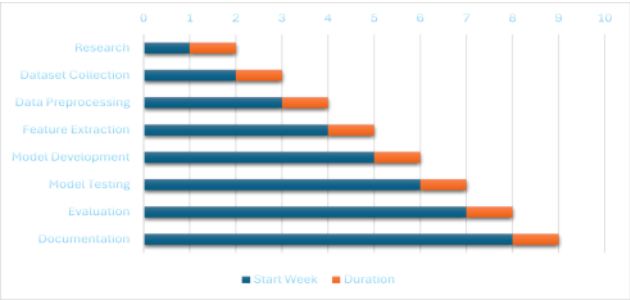
The project will follow a structured timeline to ensure that all tasks are completed efficiently. The project is expected to be completed over several stages, including research, data preparation, model development and evaluation. Initially, literature research will be conducted to understand phishing detection techniques and machine learning approaches used in cybersecurity. Following this, the dataset will be collected and analysed to identify useful features for classification.

During the development stage, machine learning algorithms such as logistic regression or decision trees will be implemented to classify emails. The model will then be trained and evaluated on test datasets to assess its accuracy and effectiveness. Finally, the results will be visualised using charts or graphs to demonstrate the phishing detection system's performance.

By defining the project scope, technologies, and timeline, this project will provide a clear plan for developing a machine-learning-based phishing-detection system to improve cybersecurity.

Gantt chart

Task	Start Week	Duration
Research	1	1
Dataset Collection	2	1
Data Preprocessing	3	1
Feature Extraction	4	1
Model Development	5	1
Model Testing	6	1
Evaluation	7	1
Documentation	8	1



Project Risks and Challenges

Every project involves potential risks and challenges that may affect the development process. Identifying these risks early allows appropriate mitigation strategies to be implemented.

One potential risk is the availability and quality of datasets. Machine learning models rely heavily on large datasets for training and testing. If the dataset contains incomplete or inaccurate information, the model's performance may be adversely affected. This risk can be reduced by using reliable datasets from reputable sources such as Kaggle or academic repositories. [9]

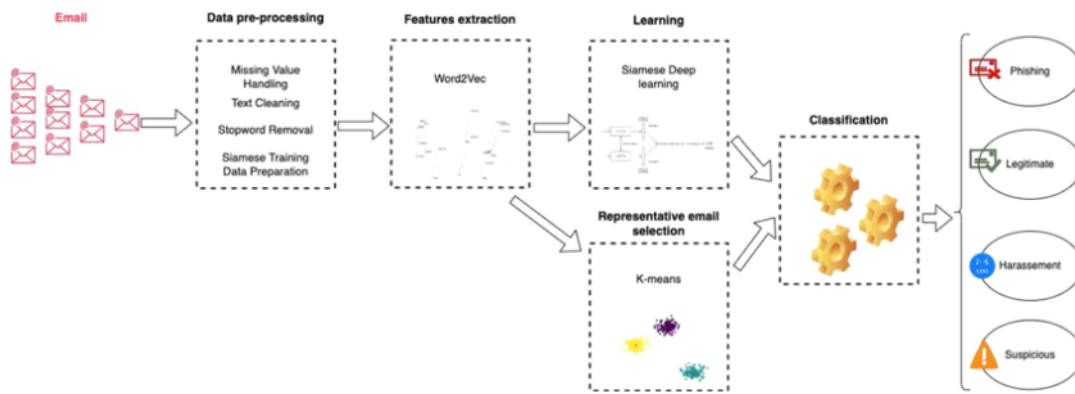
Another challenge relates to model accuracy and performance. Machine learning algorithms may initially produce low accuracy if the training data is insufficient or poorly structured. To mitigate this risk, various algorithms and feature selection techniques can be tested to improve model performance.

Technical difficulties may also arise during implementation. Issues such as software errors, coding mistakes or compatibility problems between libraries may delay the development process. Regular testing, debugging and maintaining version backups can help reduce the impact of these problems.

Time management is another important factor that could affect the project's success. If tasks are not completed according to the planned schedule, the project may fall behind. This risk can be minimised by adhering to the Gantt chart timeline and allocating sufficient time to each stage of development.

Finally, cybersecurity research often involves analysing potentially malicious data, which may pose security risks if not handled correctly. To reduce this risk, the project will be conducted within a controlled development environment, and datasets will be handled carefully.

By identifying these risks and implementing appropriate mitigation strategies, the project can be managed effectively and completed within the planned timeframe.



Task 3

04 February 2026 10:36

Development Method

A suitable development methodology for this project is Agile. Agile is widely used in software development because it allows projects to be developed incrementally through iterative cycles. Instead of building the entire system at once, the project is divided into smaller stages, allowing continuous improvements. [10]

Agile is particularly suitable for cybersecurity and machine learning projects because the development process often requires experimentation and adjustments. Machine learning models may need to be retrained several times to improve their performance. Agile allows developers to evaluate the system after each stage and make necessary improvements.

The Agile methodology typically consists of several stages, including planning, development, testing and evaluation. In the planning stage, the project requirements and objectives are defined. For this phishing detection system, the planning phase involves identifying the dataset to be used and selecting appropriate machine learning algorithms.

During the development phase, the dataset will be collected and processed using Python libraries such as Pandas and NumPy. Feature extraction techniques will be applied to convert email content into numerical data usable by the machine learning model.

The testing stage involves training the machine learning algorithm using labelled datasets and evaluating its performance using accuracy metrics. If the results are not satisfactory, model parameters can be adjusted or additional features introduced.

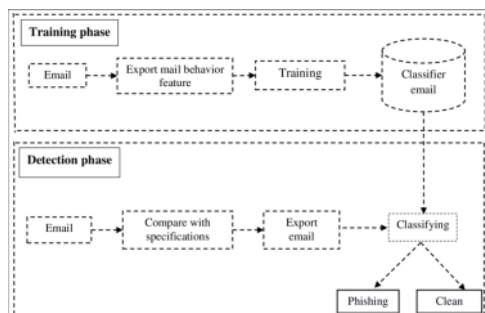
Finally, the evaluation stage assesses whether the developed system successfully detects phishing emails. The Agile approach enables continuous improvement and flexibility throughout the project lifecycle, making it well-suited to developing a phishing detection system.

Design Work

Two design elements have been created to demonstrate the structure and functionality of the proposed phishing detection system.

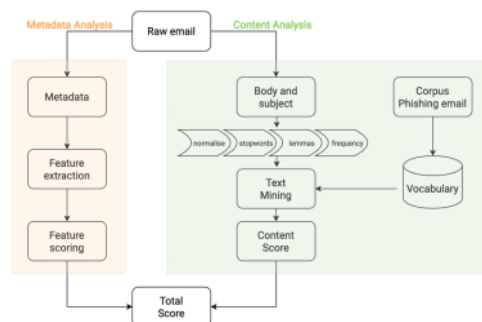
The first design element is a system architecture diagram, which illustrates the system's overall structure and how its components interact. The architecture begins with the input stage, where emails are collected. These emails are then passed to the preprocessing stage, where data cleaning and feature extraction are performed. After preprocessing, the data is analysed by a machine learning model that classifies the email as either phishing or legitimate. The final stage produces the classification output for the user.

System Architecture for the Machine Learning-Based Phishing Detection System

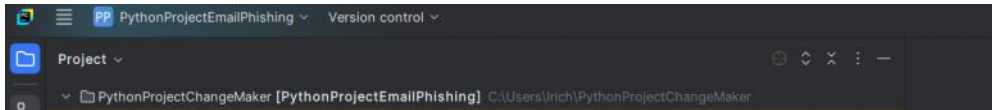


The architecture diagram illustrates how the phishing detection system processes email data. Emails are first collected and cleaned during the preprocessing stage. Relevant features are extracted from the email content before being analysed by a machine learning classification model. The system then produces a final output indicating whether the email is phishing or legitimate.

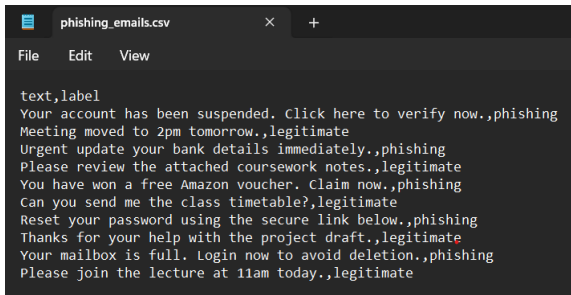
Flowchart of the Detection Process Phishing Email



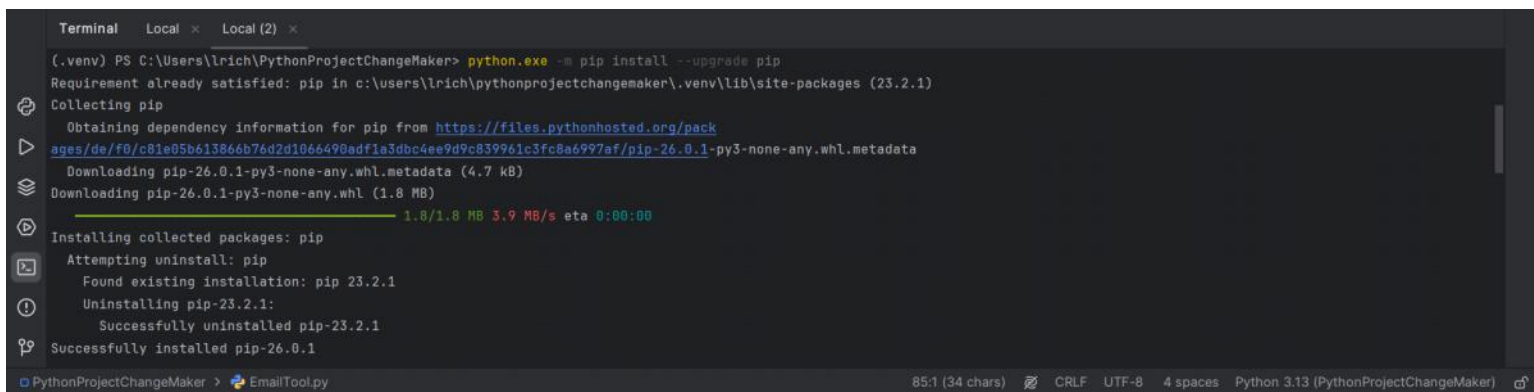
The flowchart illustrates the process for detecting phishing emails in the proposed system. When an email is received, the system extracts important features such as the sender address, subject line, attachments and message content. The email data is then pre-processed to remove unnecessary characters and prepare the text for analysis. A machine learning model analyses the extracted features to determine whether the email is malicious or legitimate. If the system identifies phishing characteristics, the email is flagged and moved to the spam or quarantine folder. Otherwise, the email is delivered to the writer's inbox safely.



First, I created a Python environment to start writing the initial code



I created a .csv file containing sample emails to train the model. The dataset used in the project contains labelled examples of phishing and legitimate emails. These labels are required during the supervised machine learning training phase to enable the algorithm to learn patterns associated with phishing behaviour. Once trained, the model can classify new unseen emails without requiring labels.



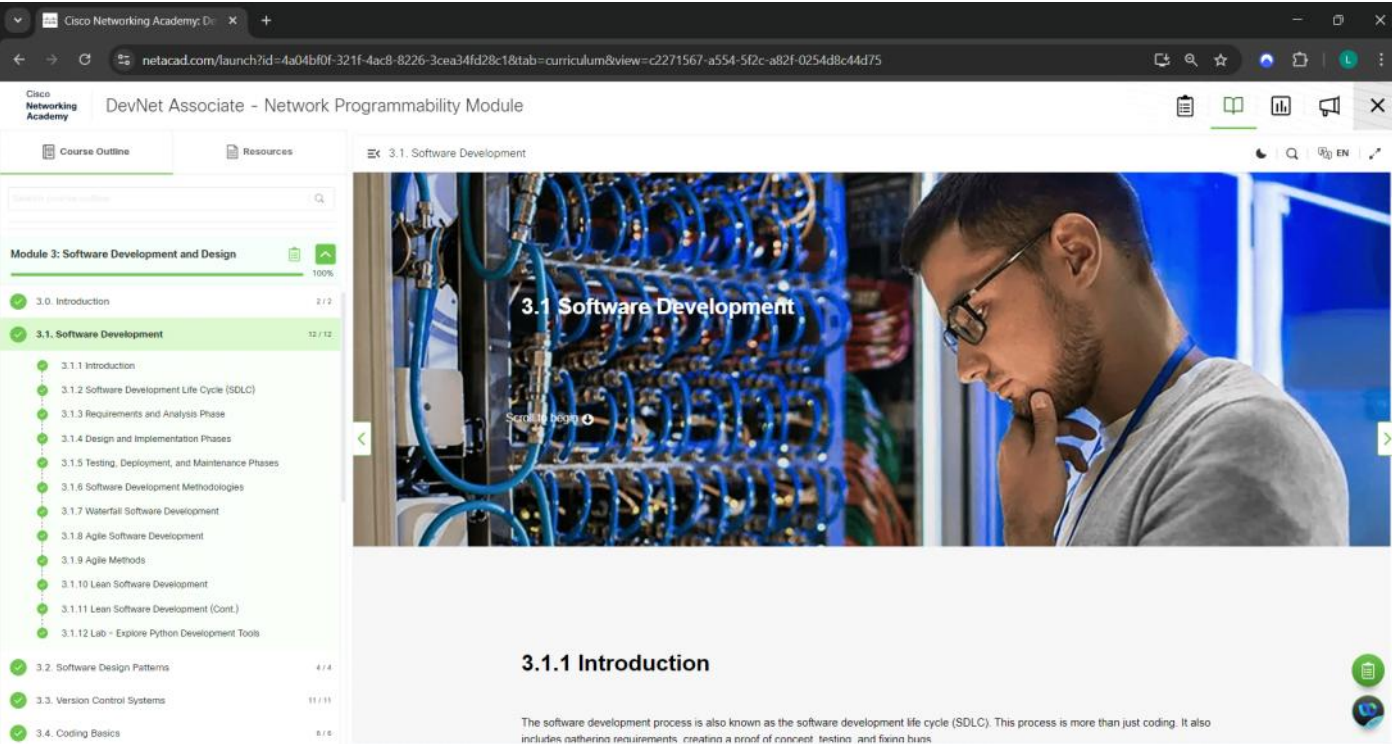
I installed some resources I needed to ensure my AI phishing tool runs correctly.

References

04 February 2026 10:37

[1]	The Blackwell Encyclopedia of Sociology, "Technological Innovation," John Wiley & Sons, Ltd., 2021.
[2]	J. C. V. C. J. S. REGNER SABILLON, "Cybercrime and Cybercriminals: A Comprehensive Study," <i>International Journal of Computer Networks and Communications Security</i> , vol. 4, no. 6, pp. 165 - 176, 2016.
[3]	J. S. a. A.-S. K. P. Kutub Thakur, "Innovations of Phishing Defense: The Mechanism,," <i>International Journal of Communication Networks and Information Security (IJCNIS)</i> , vol. 10, no. 1, pp. 19-27, 2018.
[4]	T. G. Peter Mell, "Recommendations of the National Institute," september 2011. [Online]. Available: https://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf . [Accessed march 2026].
[5]	Kaggle, "datasetd," kaggle, [Online]. Available: https://www.kaggle.com/datasets .
[6]	scikit-learn, "Machine Learning in Python," scikit-learn, [Online]. Available: https://scikit-learn.org/stable/ .

[7]	w3schools, "Pandas - Analyzing DataFrames," [Online]. Available: https://pandas.pydata.org/ .
[8]	C. S. o. Technologies, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," <i>Phishing Attacks: A Recent Comprehensive Study and a New Anatomy</i> , 09 march 2021.
[9]	A. M. &. B. S. P. S. S. Gantayat, "A Study of Incomplete Data – A Review," in <i>Springer International Publishing, Switzerland</i> , 2013.
[10]	O. S. J. R. a. J. Pekka Abrahamsson, "Agile Software Development Methods;," <i>Agile Software Development Methods;</i> , 202.



Cisco Networking Academy really helped with the whole project. Especially in this part of the network programmability module