|  |  |
| --- | --- |
| **Department of Computing and Mathematics,**  **Cyber Security Postgraduate Programmes**  **Term Of Reference Coversheet** | |
| Student Name |  |
| Student ID |  |
| Academic Supervisor |  |
| External Collaborator |  |
| Project Title | Detecting and Preventing Phishing Attacks Using Machine Learning Methods |
| Degree Title | MSc. Cyber Security |
| Project Unit Code | 6G7V0007 |
| Credit Rating | 60% |
| Start Date | 05/07/2023 |
| ToR Date | 30/06/2023 |
| Intended Submission Date | 22/09/2023 |
| Signature And Date Student |  |
| Signature And Date External Collaborator (If involved) |  |

**Contents**

Project Title......................................................................................... 3

Project Aims and Objectives.……………………….…..……………3

Learning Outcome……………………………………………………4

Project Description.............................................................................. 4

Hardware Resources Required ........................................................... 5

Software Resources Required ............................................................. 5

References .......................................................................................... 6

Evaluation plan……………………………………………………….6

Activity Schedule…………………………………………………….7

**Project Title**

Detecting and Preventing Phishing Attacks Using Machine Learning Methods

**Project Aims and Objective**

**Aims**

The aim of this project is to develop an effective model for detecting and preventing phishing attacks by leveraging the power of machine learning techniques. The project aims to enhance the security of individuals and organizations against phishing attempts, ultimately reducing the risk of financial losses, data breaches, and compromised user credentials.

**Objectives**

The objective of the project is to develop a machine learning-based system for detecting and preventing phishing attacks. It aims to create a model that accurately identifies suspicious emails and URLs in real-time. The project focuses on continuous improvement, incorporating new attack techniques and providing user awareness through proactive warnings and education. Integration with existing security infrastructure is emphasized to create a comprehensive defense against phishing. Ultimately, the project seeks to mitigate the risks associated with phishing attacks, safeguarding individuals, and organizations from financial and data loss.

**Learning Outcome**

The project's learning outcomes include gaining knowledge of phishing attacks and applying machine learning techniques for detection and prevention, I will develop skills in data analysis, model development, continuous improvement, and user awareness strategies. Integration with existing security infrastructure and understanding ethical considerations are emphasized. i will acquire the ability to assess risks and impacts, protecting individuals and organizations from financial and data loss. Overall, this project will enhance my comprehensive understanding of phishing attacks, proficiency in machine learning methods, and the ethical aspects of mitigating risks and raising user awareness.

**Project Description:**

* To develop a novel approach to detect malicious Emails and alert users. (Hany F. Atlam, 22 December 2022)
* To apply ML techniques in the proposed approach to analyze the real time URLs and produce effective results. (S. Salloum, 2022)
* developing an advanced system to identify and mitigate the threat of phishing attacks. (Ayesha Arshad, 2 Apr 2021)
* The project utilizes machine learning techniques to analyze email headers, content, URL structures, and lexical patterns to accurately detect suspicious emails and URLs in real-time. (S. Menaka, 2 February 2023).
* The project emphasizes user awareness by providing proactive warnings and education to recognize and avoid phishing attempts.
* By combining machine learning algorithms, data analysis, and user-centric approaches. (John Arthur Jupin, 1 December 2019)

**Hardware Resources Required**

**Computer System**: A laptop computer with sufficient processing power and memory capacity to support the machine learning algorithms and data processing requirements.

**Processor:** A multicore processor with a clock speed of at least 2.5 GHz or higher for efficient execution of machine learning tasks.

**Memory (RAM):** A minimum of 8 GB RAM or (16 GB or more) is desirable for handling large datasets and complex machine learning models.

**Storage:** Sufficient storage space to store the dataset, machine learning models, and any additional resources. A minimum of 100 GB.

**Network Connection:** A fast internet connection that download datasets, accessing online resources, and updating the system components as needed.

**Software Resources Required**

**Operating System**: Windows, macOS, or Linux.

**Python:** A recent version of Python programming language (e.g., Python 3.10 up) as it provides a wide range of libraries and frameworks for machine learning, such as TensorFlow, scikit-learn, and Keras.

**Integrated Development Environment (IDE):** PyCharm, Jupyter Notebook, or Anaconda can be used for developing and running the machine learning code.

**Machine Learning Libraries**: Google collab, machine learning libraries such as TensorFlow, scikit-learn, Keras, Pandas, NumPy, and Matplotlib. These libraries provide essential tools for data preprocessing, model training, and evaluation.

**Web Browser:** A web browser for accessing and analyzing phishing websites during the data collection and evaluation stages.

**Data Management Tools**: Software tools like MySQL or MongoDB can be used for storing and managing the collected dataset efficiently.

**Text Editors:** Visual Studio Code, Sublime Text, or Atom.

**Documentation Tools:** Microsoft Word

Communication and Collaboration Tools: Microsoft Teams, Slack or Google Workspace will be used for communication and collaboration with my supervisor.

# **References**

Ayesha Arshad, A. U. R. S. J. T. M. A. J. A. S. M. A., 2 Apr 2021. A Systematic Literature Review on Phishing and Anti-Phishing Techniques. *Cryptography and Security (cs.CR).*

Hany F. Atlam, O. O., 22 December 2022. *Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review.* s.l.:s.n.

John Arthur Jupin, T. S. +. a. D. S., 1 December 2019. *Review of the machine learning methods in the classification of phishing attack.* s.l.:s.n.

S. Menaka, J. H. +. a. S. M., 2 February 2023. *Analysing the Accuracy of Detecting Phishing Websites using Ensemble Methods in Machine Learning.* s.l.:s.n.

S. Salloum, T. G. +. a. K. S., 2022. *A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques.* s.l.:s.n.

**Evaluation Plan**

The purpose of this evaluation is to assess the effectiveness and performance of the machine learning-based system for detecting and preventing phishing attacks. The evaluation aims to determine the model's accuracy, robustness, and real-world applicability.

The evaluation plan for this project will involves.

* Correctly classified numbers of instances out of the total samples.
* The model's ability to process data efficiently and in a timely manner.
* Identify key activities such as data collection, model training, validation, testing, and analysis.
* Utilize a diverse dataset containing both legitimate and phishing instances for training, validation, and testing purposes.
* Assess real-world applicability, an external dataset with real phishing instances will be collected, ensuring the model is tested on unseen and practical data.
* Tools will be used to collect and maintain a database of known phishing websites to aid in training and testing and a collection of phishing emails will be compiled for studying phishing content and patterns.
* Data analysis will involve applying machine learning algorithms to the collected datasets.eg Cleaning and preparing data to handle missing values to ensure data integrity, Identifying relevant features that contribute to effective phishing detection, Training the machine learning model(s) on the training dataset using various algorithms, Assessing model performance on the testing dataset using evaluation metrics and Comparing the machine learning model(s) against baseline methods and existing solutions.

**Activity Plan**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **June** | **June** | **July** | **July** | **July** | **July** | **Aug** | **Aug** | **Aug** | **Aug** | **Sept** | **Sept** | **Sept** | **Sept** |
| 22 | 30 | 03 | 10 | 21 | 30 | 07 | 15 | 23 | 30 | 07 | 15 | 20 | 21 |
| **Tasks** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **proposal** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Project ToR** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Project Plan** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Literature Review** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Ongoing Research** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Data collection and System design** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Development** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Critical analysis and Review** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Project Submission** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

EthOs application Number 57496