

An Intelligent Phishing Detection System Using Machine Learning and Digital Forensic Indicators

Student Name: Rishabh Choudhary

Student ID: 2434332

Module: ACFA7004 – Research Methods and
Professional Development

Programme: MSc Cybersecurity and Digital
Forensics

Supervisor: Dr. Suresh Shanmugasunadaram

University: University of Wales Trinity Saint David

Submission Date: 01 May 2026

Abstract

The quick adoption and spread of digital technologies in society have led to a noticeable growth in cyber threats with phishing being one of the most frequently detected types of cybercrime. Conventional methods of phishing detections based on blacklisting and rules are now failing since they cannot cope with constantly developing and changing phishing attacks. Thus, this study focuses on creating an intelligent system for phishing detection using machine learning algorithms and digital forensics indicators.

A quantitative approach and analysis methods will be used in this study. Multiple phishing datasets available online will serve as the basis for training and testing of various machine learning models, including the Random Forest model, Support Vector Machine model, Decision Tree, and Logistic Regression. Feature-based detection based on features such as URL parameters, domain attributes, behavioral aspects will be used. Also, the digital forensics indicators will be used for improving the efficiency of detection, including metadata and activity indicators. Previously conducted research demonstrated that a combination of machine learning algorithms greatly increases the efficiency of phishing detection as opposed to conventional solutions [1]. Moreover, incorporating additional data can further increase the efficiency of detection systems [2].

The performance evaluation criteria will include classification metrics such as accuracy, precision, recall, and F1-score. The expected result is an efficient phishing detection solution able to recognize not only typical but previously unseen phishing attacks.

Table of Content

Introduction	4
LITERATURE REVIEW	5
METHODOLOGY	7
1. Research Design	7
2. Research Approach	7
3. Data Collection	7
4. Data Preprocessing and Feature Engineering	8
5. Model Development	8
6. Model Evaluation	9
7. Tools and Technologies	10
8. Ethical Considerations	10
9. System Workflow	11
10. Expected Outcomes	12
Project Timeline	12
Project Resources	13
Limitations of the Study	14
Conclusion	14

Introduction

The growing trend of technology has led to an increasing exposure of individual and organisational data towards various cybersecurity threats. One type of cybercrime which currently receives widespread attention is phishing attacks. The process involves the creation of fake e-mails, links, and websites to obtain access to sensitive data such as login credentials, passwords, etc. Due to the relatively simple nature of the crime, it is considered quite profitable among cybercriminals, leading to significant financial and reputation loss for the affected parties [2], [3].

The traditional approach to phishing detection involved the use of blacklists, rule-based filtering and other static approaches based on pre-defined malicious URL addresses. These systems proved to be ineffective since they cannot detect new phishing attempts which occur regularly due to constant changes in the phishing schemes. The use of such methods also allows attackers to create new phishing pages on regular basis. As noted in the existing literature, such limitations make it difficult to address the problem of phishing attacks in the modern environment [4].

Recently, there were several studies devoted to the introduction of innovative machine learning techniques to improve detection performance. Using algorithms of big data analysis, these tools allow for efficient identification of malicious web resources. Moreover, there are examples of classification approaches based on the analysis of website structure. Another solution includes streaming and real-time analysis to detect phishing attacks more efficiently [3]. Despite this, currently used approaches are limited to small set of features and have difficulties when dealing with complex datasets.

Another drawback in the area relates to the lack of application of digital forensics in the area. The practice involves analysis of system logs and other evidence available to help identify any suspicious activity in real time. However, there were no studies on integration of such an approach into machine learning algorithms, leading to another potential research gap.

The purpose of this research is the development of an intelligent system capable of effective phishing attacks identification. To achieve this objective, it will be necessary to analyse current detection techniques, determine their drawbacks, and propose a hybrid solution based on machine learning algorithms and digital forensics.

There are several main goals in the area. First, it will be important to consider existing detection techniques and analyse them in terms of effectiveness. Second, it

will be required to design a new phishing detection algorithm based on machine learning. Third, it will be important to test it using specific metrics, e.g. accuracy and precision. Fourth, the contribution of digital forensic factors will be analysed.

This research will contribute to the field due to its novel approach to the analysis of phishing attacks. It will allow for improved detection of cybercrimes and will help to mitigate risks associated with the issue.

LITERATURE REVIEW

Phishing detection has attracted extensive interest in the area of cybersecurity due to frequent cyberattacks. Early methods for detecting phishing websites mainly focused on utilizing heuristic and blacklisting-based algorithms. Such algorithms make use of predefined rules and databases containing malicious URLs to identify phishing websites. The method based on blacklisting compares websites against malicious domains discovered previously to block access. While such methods are highly efficient, they possess fundamental disadvantages in terms of detecting novel phishing websites because they cannot adapt to new attacks [4].

The problem of static phishing detection led researchers to investigate other ways to identify malicious sites. Some studies showed that it was possible to detect phishing using the analysis of website structural parameters. The analysis found the following features that could help to separate phishing URLs from benign ones: length of a URL, age of a domain, and suspicious symbols in it [3]. client-side algorithms that were trained to recognise such features using a machine learning technique. Such a development is a considerable progress in phishing detection since the model can generalize the knowledge about the attacks and therefore can deal with novel phishing attacks better than previous ones [2]. Nevertheless, the performance of these methods depends strongly on the selection of features used for training.

Machine learning became the dominant way of tackling the phishing problem. Various classification algorithms such as Support Vector Machine, Decision Tree, and Random Forest have been extensively used to detect phishing attacks. A study proposed a dynamic phishing detection system based on fuzzy data mining, which proved to be a highly adaptive solution [1]. Besides, the latest frameworks include the use of streaming analytics for detecting phishing attacks in real-time [5]. All these advances prove the effectiveness of machine learning approaches to detecting threats.

Nevertheless, the use of machine learning faces various challenges in this area. Firstly, machine learning models suffer from generalisation problems, which means that a model trained with a particular dataset may fail to detect threats outside of the range defined by this dataset. Secondly, much research uses static datasets and thus do not reflect the dynamical nature of threats. It leads to good results during training but poor performances in the real-world scenarios. Thirdly, most of the existing approaches focus only on the URL features or web page content and ignore other parameters of the site.

It should be noted that despite the prevalence of machine learning approaches, their integration with digital forensics is under-researched. It is widely known that digital forensics deals with the analysis of the attack trace, usually represented by system and network logs. Digital forensic methods play an essential role in threat detection and can provide useful indicators that may assist in creating a detection model. However, most of current models ignore any data related to digital forensics and operate independently [6].

Another challenging point that arises from modern research is the sophistication of phishing attacks. Phishing attackers tend to apply more complicated obfuscation methods as well as use domain generation algorithms to prevent recognition. As a result, detection systems must evolve to detect those attacks successfully. Machine learning provides researchers with a convenient base that can be further developed to meet the above-mentioned requirements.

To summarise, there is several issues regarding the current research in the field. Firstly, the models trained via machine learning algorithms fail to generalize the knowledge in some cases. Secondly, static datasets are widely used despite their inefficiency for detecting modern threats. Thirdly, existing approaches mostly concentrate on a limited number of features to train algorithms. Fourthly, integration with digital forensics is hardly ever used. Based on the above, there appears a gap in current research in terms of creating a detection model that would combine machine learning and digital forensics.

METHODOLOGY

1. Research Design

For this project, I will adopt a quantitative research design that will employ experiments to create and test the proposed system for intelligent phishing detection. The research will be mainly centred around the analysis of structured datasets of phishing and non-phishing (legitimate) websites with the help of machine learning methods. Quantitative research design will be appropriate for this case, because its purpose is measuring how well the system works (accuracy, recall rate, etc.).

As for the specifics of my experiment, it implies testing of various machine learning algorithms (and even combinations thereof) for determining which one can detect phishing activities effectively. In addition, the proposed model will include digital forensic indicators for checking how important it is for the overall effectiveness. In other words, the experiment will be performed systematically and will allow for reproducibility.

2. Research Approach

This particular research will rely on the data-driven approach, whereby knowledge is obtained through systematic analysis of data. It consists of collecting data, processing it, extracting key features, building models based on them, and evaluating model performance.

It is supposed to conduct supervised machine learning as there are some labelled datasets, so I can build classifiers for detecting phishing cases.

3. Data Collection

This research is going to make use of available phishing databases in order to eliminate ethical concerns. Possible sources are listed below:

Available phishing website databases from various websites such as Kaggle.
Datasets from UCI Machine Learning Repository.
Phishing open-source repositories providing phishing URLs.

All datasets include the following features:

- Various URL attributes (length, number of special characters).
 - Different aspects concerning domains (age, DNS records).
 - Website behaviour parameters.
-
- Ethically speaking, all data will be available online, so there will be no risks connected to any kind of sensitive information.

4. Data Preprocessing and Feature Engineering

Data preprocessing and feature engineering are necessary steps before implementing the models, which consist of the following tasks:

Removal of duplicates and irrelevant instances;
 Handling missing values, if any;
 Normalizing/scale numerical values;
 Encoding categorical variables;

Feature engineering refers to creating additional features by analysing URL or website content. For example, this could include checking how complex URL structure looks and searching for suspicious patterns.

In addition, digital forensic data will be incorporated where possible (for instance, some metadata features or behavioural patterns).

5. Model Development

As part project, several machine learning algorithms will be implemented and tested for effectiveness in detecting phishing activities. Some possible algorithms are listed below:

- Logistic regression.
- Decision tree.
- Random forest.
- Support vector machine (SVM).

Many different machine learning algorithms were analyzed for potential inclusion in the experiment. The comparison of several selected algorithms together with their advantages and disadvantages is provided in Table

Model	Strengths	Limitations
Logistic Regression	Simple, fast, interpretable	Limited for complex patterns
Decision Tree	Easy to understand, visualisable	Prone to overfitting
Random Forest	High accuracy, robust	Computationally expensive
SVM	Effective in high dimensions	Slower on large datasets

All models will be trained on the pre-processed dataset and optimized (wherever possible). After that, all the results will be compared in terms of effectiveness and the best one chosen.

Another task will be developing a hybrid model with the incorporation of digital forensics features and comparing its performance with the previously selected one.

6. Model Evaluation

Evaluation of the models developed will involve the use of standard classification measures, namely:

- Accuracy.
- Precision.
- Recall.
- F1-score.

The quality of the constructed algorithms will be estimated according to commonly used measures for classification tasks. The description of selected measures together with their roles is provided in Table

Metric	Description
Accuracy	Measures overall correctness of the model
Precision	Indicates proportion of true positives among predicted positives
Recall	Measures ability to detect phishing instances
F1-score	Harmonic mean of precision and recall

These criteria will ensure that the selected algorithm is effective, especially when dealing with unbalanced datasets (when phishing instances do not occur too frequently).

In addition, I will use confusion matrices to evaluate the model in terms of accuracy.

7. Tools and Technologies

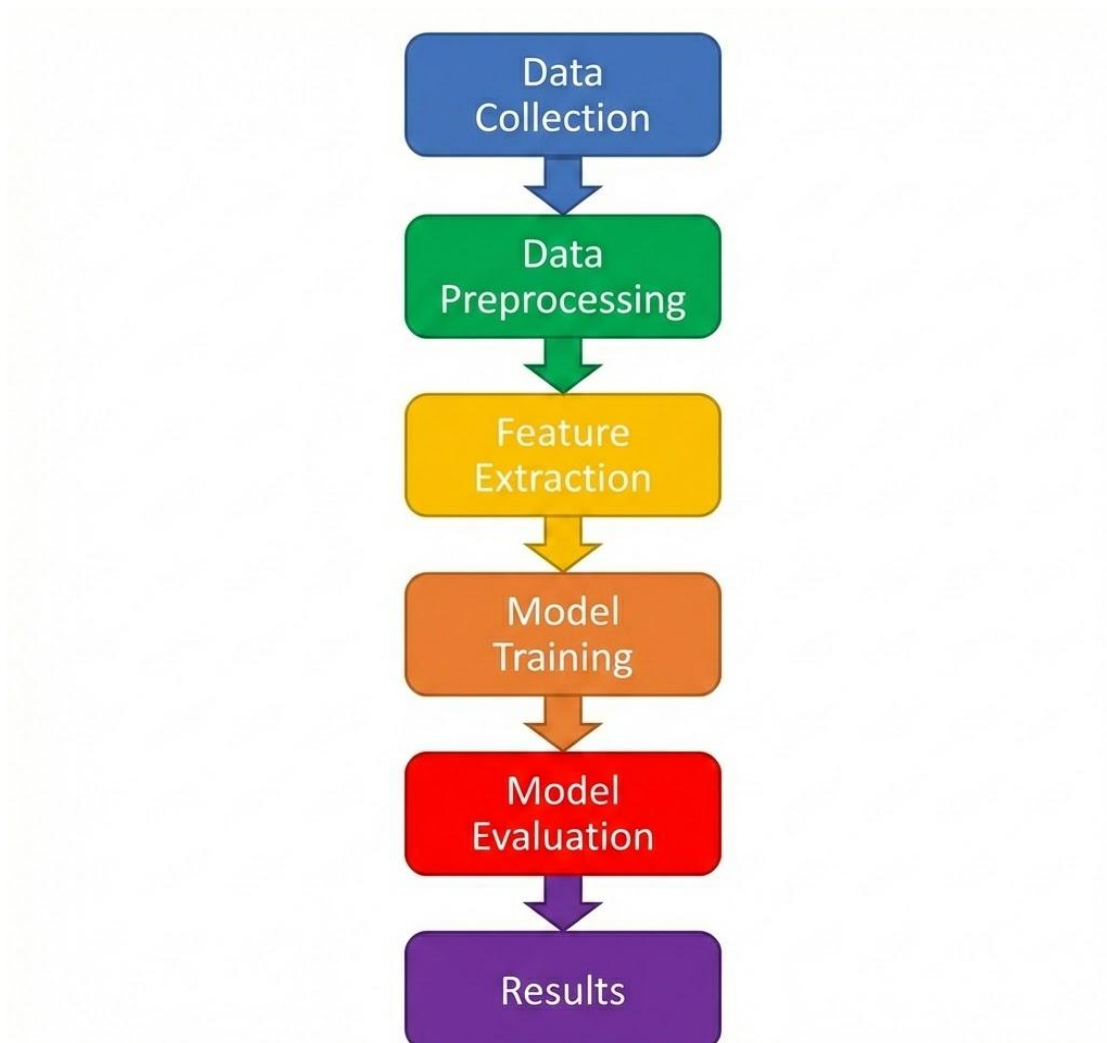
I will use the following technologies as part of this project:

- Python.
- Scikit-learn, Pandas, NumPy.
- Google Colab or Jupyter notebook.

8. Ethical Considerations

This research makes use of datasets available online and does not involve any humans (or personal information whatsoever). As far as ethical considerations go, there will be no risks at all.

9. System Workflow



10. Expected Outcomes

The principal expected outcome of this research includes developing a smart phishing detection system with capabilities of distinguishing phishing websites from genuine sites. It is expected that machine learning-based approaches will outperform conventional rule-based detection systems especially in detecting new or zero-day attacks. This prediction aligns with previous research findings that data-driven approaches exhibit better adaptability in evolving cyberattacks [4].

The inclusion of forensic evidence will improve the effectiveness of the detection framework by leveraging additional contextual information like metadata and behaviour analysis. As a result, the hybrid detection system is expected to achieve higher precision rates with reduced false positives. This finding is crucial in improving the efficiency of cybersecurity operations and enhancing trust in the detection systems.

Moreover, the research will provide an evaluation and comparison of different machine learning models and their appropriateness in detecting phishing attacks. Such findings will identify the suitable algorithm for certain types of phishing attacks.

Therefore, the outcomes of this research will advance current cybersecurity practices by designing a more adaptive and efficient approach to detecting phishing attacks. The proposed framework can provide a basis for future studies to integrate machine learning and digital forensics in cybersecurity.

Project Timeline

The research will be implemented within a systematic timeline that enables efficient execution and successful completion within the set period. The project is structured into different stages focused on accomplishing various phases of the research.

During the first phase, attention will be given to reviewing literature sources and formulating the research problem. The second stage will focus on collecting and processing relevant datasets for the analysis. In the third stage, the design and training of machine learning models will be accomplished. Subsequently, the fourth phase will involve evaluating the models' performance based on predetermined criteria. Finally, the project will culminate in report writing and final review.

Phase	Activity	Duration
Phase 1	Literature Review and Problem Definition	Weeks 1-3
Phase 2	Data Collection and Preprocessing	Weeks 4-6
Phase 3	Model Development and Training	Weeks 7-9
Phase 4	Model Evaluation and Analysis	Weeks 10-11
Phase 5	Report Writing and Final Review	Weeks 12-13

The outlined timeline will enable efficient utilisation of time and resources while leaving room for refinement during implementation.

Project Resources

The successful implementation of this research necessitates access to adequate resources. In particular, the research will require computing resources to perform data analysis and develop machine learning models. Considering the relatively low computational complexity of the selected models, ordinary computing environments will suffice.

The project will leverage the Python programming language with popular machine learning libraries such as Scikit-learn, Pandas, and NumPy. These tools will facilitate efficient model development, analysis, and evaluation. Development platforms such as Jupyter Notebook or Google Colab will be used to carry out experiments.

Apart from computing resources, the study will need access to publicly available phishing datasets and academic research databases to support both experimental and theoretical research aspects.

Limitations of the Study

Even with an elaborate methodology, several limitations can affect the outcomes of this research. Firstly, the use of publicly available datasets might constrain the generality of the developed models due to their lack of comprehensive coverage of phishing attacks. Consequently, the models' application in live environments will be limited.

Additionally, the effectiveness of machine learning models will depend significantly on the representation of relevant features. The poor quality and biased representation of features may impair model performance. However, measures will be taken to incorporate all necessary features, including forensic indicators. Nevertheless, the availability of forensic evidence in the datasets will pose a challenge.

The dynamic nature of phishing attacks is another factor that might limit the performance of static models. Although machine learning models offer flexibility, periodic re-training and update of models will be necessary in live environments.

Finally, the scope of the research will be restricted due to limited time and computing resources, which cannot allow for the investigation of advanced approaches such as deep learning.

These limitations provide potential avenues for future research and development.

Conclusion

To conclude, this study explores the design of an intelligent phishing detection system that uses machine learning models in combination with digital forensic indicators. The research aims to address key shortcomings found in existing literature including the incapacity of traditional approaches to detecting evolving phishing attacks, inadequate generalisation capabilities of machine learning models, and absence of forensic data in detection frameworks.

With a data-driven methodology, the research evaluates the effectiveness of several machine learning models and examines the impact of including forensic data in the detection process. The outcomes are expected to exhibit improved detection accuracy and robustness, hence making meaningful contributions to the development of reliable cybersecurity solutions.

Overall, the proposed framework presents a comprehensive solution to phishing detection and paves the way for further studies in integrating artificial intelligence and digital forensics in cybersecurity applications.

References

- [1] A. K. Jain and B. B. Gupta, "Towards detection of phishing websites on client-side using machine learning based approach," vol. 68, no. 4, pp.687–700, 2018,
- [2] R. Mohammad, F. Thabtah and L. McCluskey, "Phishing websites features: Classification and detection," vol. 6, no. 1, pp.1–19, 2015,
- [3] S. Marchal, J. François, R. State and T. Engel, "PhishStorm: Detecting phishing with streaming analytics," vol. 11, no. 4, pp.458–471, 2014,
- [4] M. Aburrous, M. A. Hossain, F. Thabtah and K. Dahal, "A dynamic phishing detection system using fuzzy data mining," *Expert Syst.Appl.* vol. 38, no. 5, pp.5896–5906, 2011,
- [5] J. Ma, et al., "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs",
- [6] E. Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet*, Academic Press, 2011.