

INTELLIGENT PHISHING DETECTION SYSTEM

USING MACHINE LEARNING & DIGITAL FORENSIC INDICATORS

Student Name: Rishabh Choudhary

Student Id: 2434332

Module: ACFA7004

MSc Cybersecurity & Digital Forensics

REASERCH CONTEXT

- Rapid development of digital technologies has led to increasing cyber threats
- Phishing is one of the most common cyberattacks
- It is used for stealing sensitive data (e.g. passwords, bank data etc.)
- It causes great financial and reputational damages
- Machine Learning is commonly used to detect such threats [1], [2]

But phishing is still a major problem in cybersecurity.

BACKGROUND

Traditional methods for detecting phishing attacks include:

- Blacklisting
- Rule-based filtering

More advanced methods such as:

- Using Machine Learning algorithms
- Creating systems for real-time detection [4]
- Using feature-based methods

But there are still many difficulties in detecting phishing attacks

RESEARCH PROBLEM

The conventional techniques rely on signatures of known attacks.

But they cannot:

- Detect phishing attacks that are zero-days.
- Detect emerging types of phishing scams.
- Such methods can be easily exploited by the attacker.

There is a need for adaptable and real-time detection systems [3].

RESEARCH GAP

- Poor generalization capabilities of ML models
- Usage of fixed data sets
- Limited attention to input data (URLs only)
- Not enough usage of forensic information.

Need for hybrid method using machine learning and forensic information.

AIMS & OBJECTIVES

Research Aim:

Develop an intelligent phishing detection system based on machine learning and forensic indicators.

Objectives:

- To review existing techniques for detecting phishing.
- Identify their weaknesses.
- Create ML-based phishing detection model.
- Test model with static measures of its effectiveness (accuracy, precision, recall).

LITERATURE REVIEW SUMMARY

- ML improves phishing detection effectiveness [1], [2].
- Feature-based approaches are popular.
- Proposed a new real-time detection framework [3].
- Traditional techniques are inefficient against new attacks [4].

Gap:

- Limited usage of forensic indicators.

RESEARCH METHODOLOGY

- Experimental methodology
- Data driven approach
- Supervised Machine learning algorithms

Models Used:

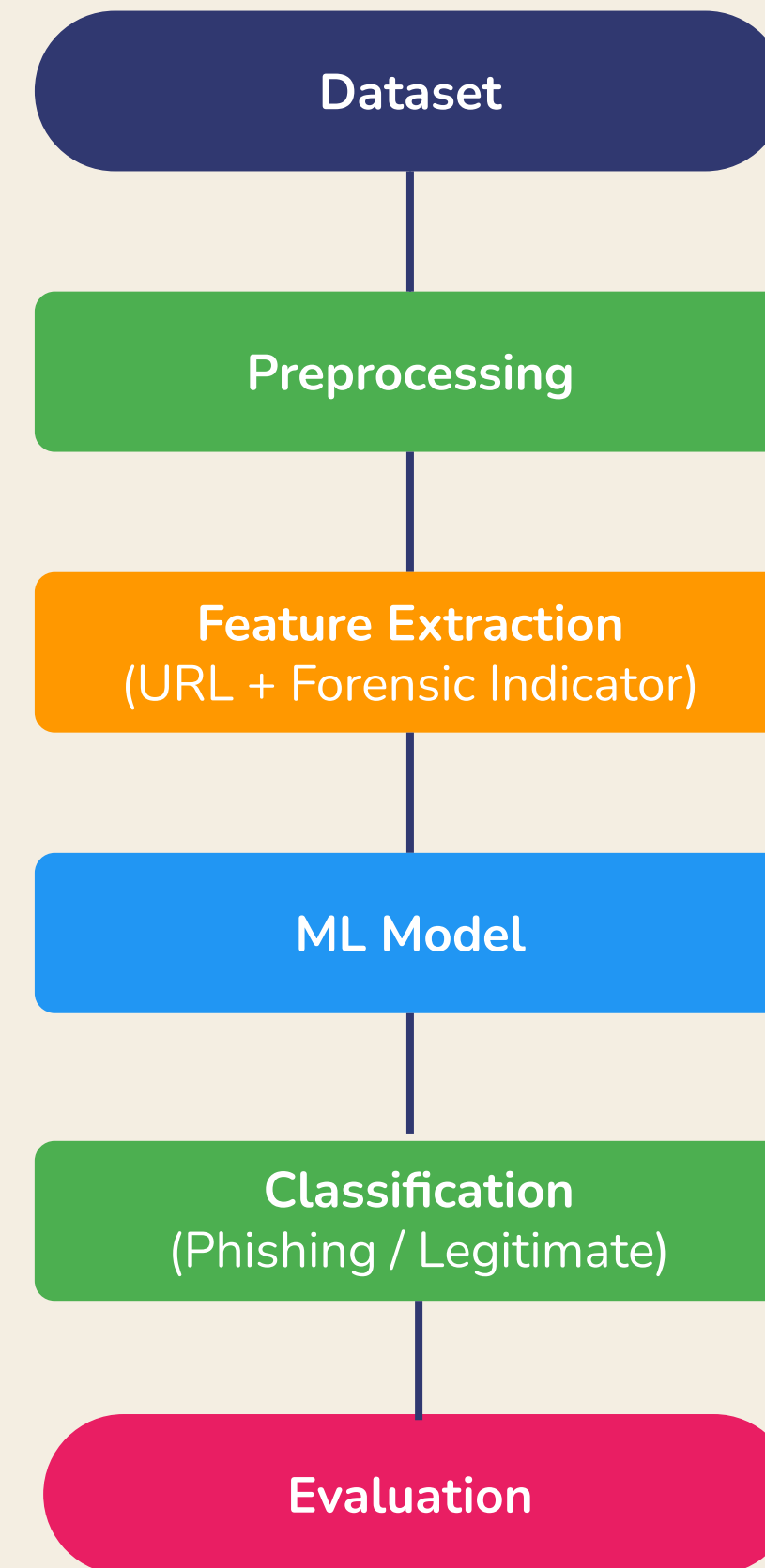
- Random Forest
- Logistic Regression
- Decision Trees
- SVM

Using a quantitative methodology to objectively evaluate model performances.

METHODOLOGY PROCESS

Key points:

- Use phishing data sets (from Kaggle/UCI)
- Extract features from them (URL + Domain + forensic information).
- Train models and test on a set



MODEL EVALUATION

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

A confusion matrix will be used to analyse classification result.

- *Ensures reliable model comparison.*

PROJECT PLANNING & RESOURCES

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

Resources Required:

- Python
- Scikit-learn
- Google Colab

IMPACT & LIMITATIONS

Impact:

- Improved effectiveness of phishing detection.
- Improved cybersecurity systems.
- Reducing risks of financial damages.

Limitations:

- Datasets constraints.
- Generalisation problems with models.
- Insufficient forensic information.

CONCLUSION

- The problem of phishing is still relevant today.
- Conventional approaches are not adequate for addressing new threats.
- Application of machine learning techniques enhances detection.
- Inclusion of digital forensics features further improves system efficiency.
- The proposed technique offers an improved method of detection.
- This study will advance the field of cybersecurity.

REFERENCE

[1] A. K. Jain and B. B. Gupta,

"Towards detection of phishing websites on client-side using machine learning based approach,"

Telecommunication Systems, vol. 68, no. 4, pp. 687-700, 2018.

[2] R. Mohammad, F. Thabtah and L. McCluskey,

"Phishing websites features: Classification and detection",

Journal of Information Security, vol. 6, no. 1, pp. 1-19, 2015.

[3] S. Marchal, J. François, R. State and T. Engel,

"PhishStorm: Detecting phishing with streaming analytics"

IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458-471, 2014.

[4] J. Ma, L. K. Saul, S. Savage and G. M. Voelker,

"Beyond blacklists: Learning to detect malicious web sites from suspicious URLs,"

Proc. ACM SIGKDD, 2009.

THANK YOU

Presented by Rishabh Choudhary

Module: Research Methods and Professional
Development

Lecturer: Dr. Suresh Shanmugasunadaram