Symbiosis Institute of Technology, Pune

Symbiosis International (Deemed University)

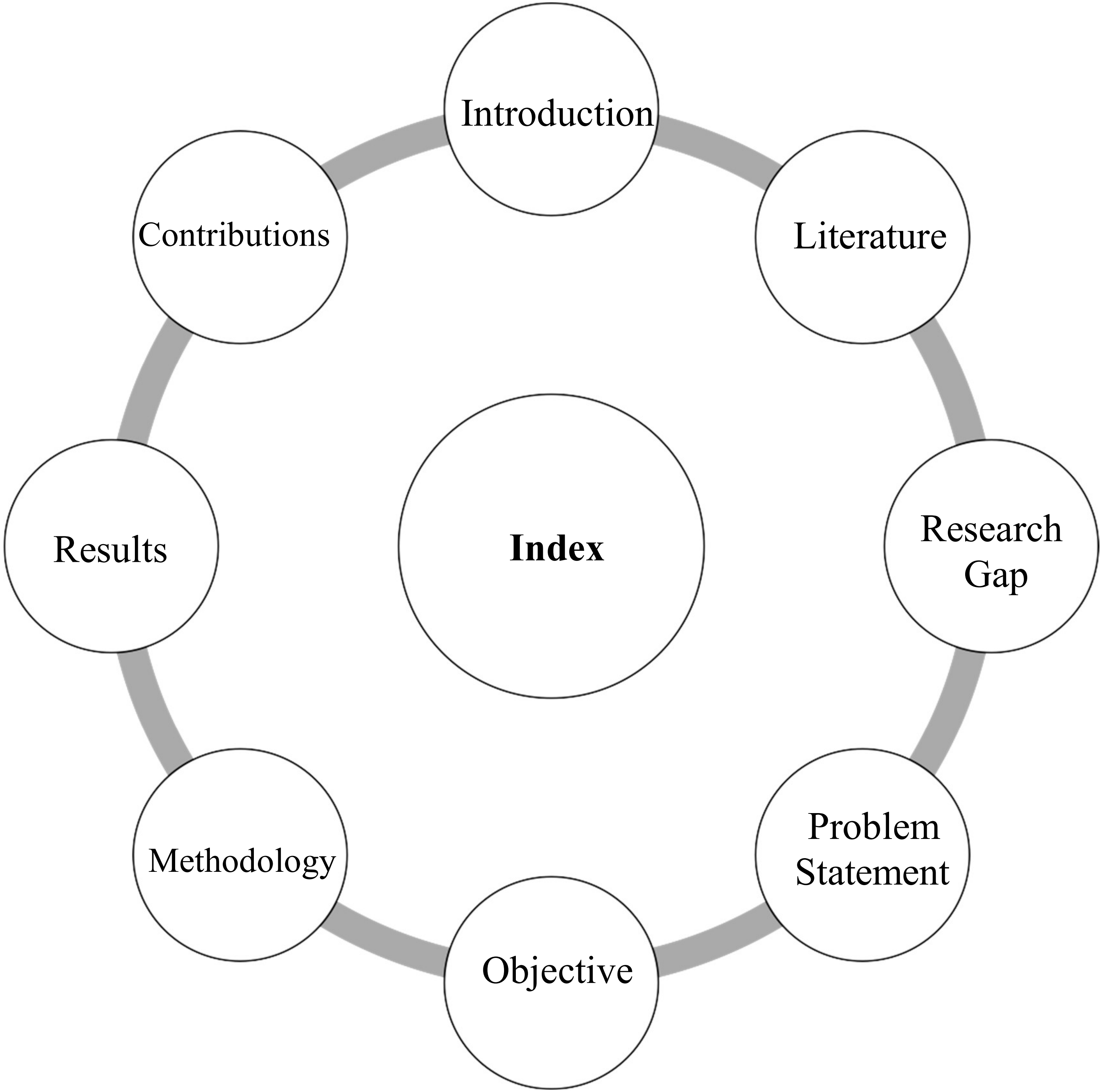
Department of Robotics & Automation Engineering

Web Phishing Detection using Machine Learning

Tuesday, September 3, 2024

Presented by

1. Ansh Sharma (22070127011)
2. Priyanshu Lathi (22070127048)



Index

Introduction

Literature

Research Gap

Problem Statement

Objective

Methodology

Results

Contributions

Introduction

Web phishing is a fraudulent tactic where cybercriminals impersonate legitimate organizations to deceive individuals into sharing sensitive information like passwords or credit card details. They typically use fake emails or websites that closely resemble the real ones. The aim is to trick users into divulging personal data, which can be used for identity theft or financial fraud. Phishing attacks can also occur through text messages or phone calls. Protecting against phishing involves being cautious with electronic communications, verifying the authenticity of websites and requests, and employing security measures like email filtering and multi-factor authentication to mitigate risks.

Literature Survey

Research Papers	Methodology	Feature extraction technique	Model/ algorithm	Accuracy
Phishing Website Detection Using Machine Learning: A Review By - Dr. Nahla Abbas Flayh And Marwa Abd Al Hussein Qasim	Feature extraction, using different models to train and detecting malicious websites in real time		Decision tree, Random Forest and Support vector machine	93%,97%,96%
Detecting Phishing Websites Using Machine Learning By- Aniket Garje1 , Namrata Tanwani1 , Sammed Kandale1 , Twinkle Zope1 , Prof. Sandeep Gore2	Feature extraction,model training,cross validation and detecting malicious websites in real time		Decision Tree,KNN, Naïve byes,	99%,97%,96%

Research Papers	Methodology	Feature extraction technique	Model/ algorithm	Accuracy
Phishing Website Detection using Machine Learning Algorithm	Feature Extraction,splilting the data, training and validation		Decision Tree, Random Forest, Support vector machine	97.14%

Research Gap:

- Get better accuracy than existing models
- Reduce computational time
- Detecting malicious websites in real time

Problem Statement

Develop a machine learning-based solution for detecting web phishing attempts in real-time. The objective is to create a robust system capable of accurately identifying fraudulent websites that impersonate legitimate entities to steal sensitive information from users. The solution should leverage features such as website content, URL characteristics, and server information to train and deploy a model capable of distinguishing between genuine and phishing websites. The system should be scalable, efficient, and able to adapt to emerging phishing tactics. Evaluation metrics should include accuracy, precision, recall, and F1-score, with a focus on minimizing false positives to enhance user trust and security.

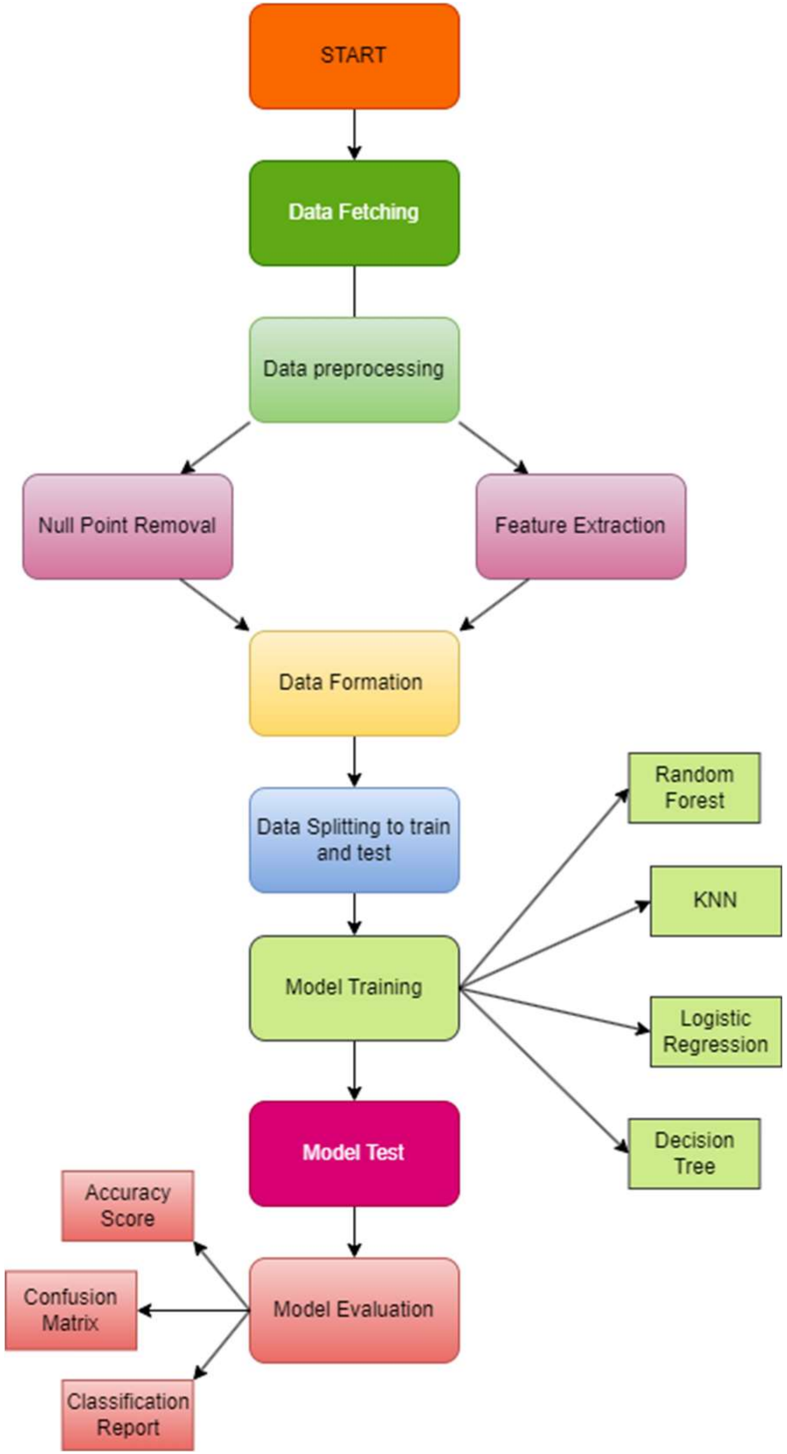
Objective:

1. Identify and classify fraudulent websites accurately: Train a model to differentiate between legitimate and phishing websites by analyzing various features such as website content, URL characteristics, and server information.

2. Ensure real-time detection: Implement a solution capable of quickly analyzing web pages as users access them, enabling prompt detection and prevention of phishing attempts..

3. Enhance security and user trust: Minimize false positives and false negatives to improve the system's reliability and user confidence in detecting phishing attacks.

Methodology and Method



Results and discussion

Following results were recorded after model evaluation:

Model	Accuracy
Random Forest	97%
KNN	95%
Logistic Regression	92%
Decision Tree	96%

Conclusion

In conclusion, this project has demonstrated the effectiveness of machine learning algorithms in detecting web phishing attempts using a comprehensive dataset of URL and webpage attributes. The high accuracies achieved by K-Nearest Neighbors, Logistic Regression, and Random Forest models emphasize the potential of these techniques in bolstering cybersecurity measures against phishing attacks.

The study underscores the importance of feature selection and dataset quality in training robust phishing detection models. By leveraging diverse features such as URL structure, webpage content, and domain characteristics, the models were able to discern subtle indicators of phishing attempts with notable accuracy.

Moving forward, further advancements in machine learning techniques, coupled with continuous updates to datasets reflecting evolving phishing tactics, are essential to stay ahead of cyber threats. Additionally, the integration of real-time monitoring and adaptive learning mechanisms could enhance the agility and responsiveness of phishing detection systems.

Overall, this research contributes to the ongoing efforts in cybersecurity by providing insights into the efficacy of machine learning in combating web-based phishing attacks. By leveraging the power of data-driven approaches, organizations can fortify their defences and mitigate the risks posed by malicious actors in the digital landscape.