DETECÇÃO DE WEBSITES DE PHISHING UTILIZANDO MACHINE LEARNING: uma análise comparativa de algoritmos de classificação

PHISHING WEBSITES DETECTION USING MACHINE LEARNING: a comparative analysis of classification algorithms

José Mirosmar
jmss6@discente.ifpe.edu.br
João Almeida e Silva
joao.almeida@belojardim.ifpe.edu.br

RESUMO

O crescente número de ameaças cibernéticas, com destaque para os ataques de phishing, representa um risco significativo para a segurança de dados de usuários e empresas. Este trabalho desenvolveu e avaliou modelos de Machine Learning como uma abordagem para a detecção automática e inteligente de websites de phishing. O objetivo foi comparar a eficácia dos algoritmos de classificação Random Forest e Support Vector Machine (SVM) na identificação de URLs maliciosas. A metodologia foi baseada em uma abordagem quantitativa, utilizando um dataset público da plataforma Kaggle. Os dados foram pré-processados e analisados, servindo de base para o treinamento e teste dos modelos. Os resultados demonstraram que o modelo Random Forest alcançou uma acurácia de 98.20%, superando significativamente o desempenho do SVM, que obteve 86.35%. Concluiu-se que, para o conjunto de dados e as condições avaliadas, o Random Forest é uma abordagem mais robusta e eficaz para o problema proposto, contribuindo para o avanço das pesquisas em segurança da informação.

Palavras-chave: Segurança da Informação. Phishing. Machine Learning. Random Forest.

ABSTRACT

The increasing number of cybersecurity threats, especially phishing attacks, poses a significant risk to the data security of users and companies. This work developed and evaluated Machine Learning models as an approach for the automatic and intelligent detection of phishing websites. The objective was to compare the effectiveness of the Random Forest and Support Vector Machine (SVM) classification algorithms in identifying malicious URLs. The methodology was based on a quantitative approach, using a public dataset from the Kaggle platform. The data was pre-processed and analyzed to serve as a basis for training and testing the models. The results demonstrated that the Random

Instituto Federal de Pernambuco. Campus Belo Jardim. Curso de Bacharelado em Engenharia de Software. 14 de setembro de 2025.

Forest model achieved an accuracy of 98.20%, significantly outperforming the SVM model, which obtained 86.35%. It was concluded that, for the dataset and conditions evaluated, Random Forest is a more robust and effective approach for the proposed problem, contributing to the advancement of research in information security.

Keywords: Information Security. Phishing. Machine Learning. Random Forest.

---

1 INTRODUÇÃO

A onipresença da internet transformou a sociedade, mas também introduziu novas vulnerabilidades. Dentre as ameaças cibernéticas, o *phishing* se destaca como um dos ataques mais prevalentes e danosos, visando enganar usuários para que revelem informações sensíveis, como credenciais de acesso e dados financeiros. A crescente sofisticação destes ataques torna a detecção manual insuficiente, criando uma demanda por soluções automáticas e inteligentes.

Neste contexto, o Machine Learning (ML) surge como uma abordagem promissora (**hastie2009elements**). Ao treinar algoritmos para reconhecerem os padrões característicos de URLs maliciosas, é possível desenvolver ferramentas capazes de identificar ameaças em tempo real com alta precisão, como demonstrado em estudos recentes na área (**mandadi2022**).

1.1 Justificativa

A relevância deste trabalho reside no potencial de mitigar os impactos negativos do phishing. A criação de modelos de detecção eficazes contribui diretamente para a segurança de usuários, a proteção da reputação de empresas e a redução de perdas financeiras. Academicamente, a pesquisa avança o estado da arte ao comparar algoritmos específicos para este problema, gerando conhecimento aplicável tanto no meio industrial quanto no científico.

1.2 Problema de Pesquisa

A questão central que norteia este trabalho é: De que forma e com qual eficácia os algoritmos de Machine Learning, especificamente Random Forest e Support Vector Machine, podem ser aplicados para a detecção automática de websites de phishing com base em características de suas URLs?

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver e avaliar a performance de modelos de Machine Learning para a detecção de websites de phishing.

1.3.2 Objetivos Específicos

- Realizar uma revisão bibliográfica sobre os temas de phishing e algoritmos de classificação;
- Analisar e pré-processar um conjunto de dados públicos sobre o tema;
- Implementar e treinar os modelos de classificação Random Forest e Support Vector Machine;
- Comparar o desempenho dos modelos utilizando métricas de avaliação como acurácia, precisão e recall.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Phishing: Conceitos e Técnicas

2.2 Machine Learning e Aprendizado Supervisionado

2.3 Algoritmos de Classificação

2.4 Trabalhos Correlatos

3 METODOLOGIA

Este trabalho seguiu uma abordagem de pesquisa quantitativa e experimental. A metodologia foi dividida nas seguintes etapas:

1. **Fonte de Dados:** Foi utilizado o dataset público “Phishing Dataset for Machine Learning” (**kaggle_dataset_2022**), cuja base original foi apresentada por **mandadi2022**. Este conjunto de dados contém um vasto número de amostras e características já extraídas de URLs, as quais são previamente classificadas como phishing ou legítimas.
2. **Ferramentas:** O desenvolvimento foi realizado na linguagem Python, com o auxílio das bibliotecas Pandas para manipulação de dados e Scikit-learn (**scikit-learn**) para a implementação dos modelos de Machine Learning, além de Matplotlib/Seaborn para a visualização de dados.
3. **Tratamento dos Dados:** Foi realizada uma análise exploratória para compreender a distribuição e correlação dos dados. Posteriormente, o conjunto de dados foi dividido em 80% para o conjunto de treino e 20% para o conjunto de teste, utilizando a função *train_test_split* da biblioteca Scikit-learn.
4. **Modelagem e Avaliação:** Foram treinados e avaliados os algoritmos Random Forest e Support Vector Machine. A performance dos modelos foi comparada utilizando as métricas de Acurácia, Matriz de Confusão, Precisão e Recall.

4 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados e analisados os resultados obtidos a partir da execução da metodologia descrita.

4.1 Desempenho dos Modelos

Tabela 1 – Resultados comparativos de desempenho dos modelos.

Métrica	Random Forest	SVM
Acurácia	98.20%	86.35%
Precisão (Phishing)	0.98	0.90
Recall (Phishing)	0.98	0.82
Falsos Positivos	18	181

Fonte: O autor (2025)

4.2 Análise Comparativa e Discussão

Conforme apresentado na Tabela **??**, o modelo Random Forest demonstrou uma superioridade significativa em todas as métricas avaliadas.

5 CONCLUSÃO

5.1 Síntese dos Resultados

5.2 Limitações do Trabalho

5.3 Trabalhos Futuros