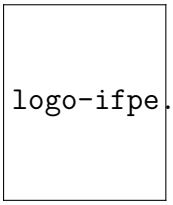

DETECÇÃO DE WEBSITES DE PHISHING UTILIZANDO MACHINE LEARNING: uma análise comparativa de algoritmos de classificação

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 propõe o desenvolvimento e a avaliação de modelos de Machine Learning como uma abordagem para a detecção automática e inteligente de websites de phishing. O objetivo geral é comparar a eficácia dos algoritmos de classificação Random Forest e Support Vector Machine (SVM) na identificação de URLs maliciosas. A metodologia será baseada em uma abordagem quantitativa, utilizando um dataset público da plataforma Kaggle. Os dados serão pré-processados e analisados, para então servirem de base para o treinamento e teste dos modelos. A avaliação será conduzida por meio de métricas de desempenho como acurácia, precisão e recall. Espera-se, ao final, determinar qual dos modelos apresenta maior eficácia para o problema proposto, contribuindo para o avanço das pesquisas em segurança da informação.

Palavras-chave: Segurança da Informação. Phishing. Machine Learning. Classificação de Dados.

ABSTRACT

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

Instituto Federal de Pernambuco. Campus Belo Jardim. Curso de Bacharelado em Engenharia de Software. 11 de julho de 2025.

models. The evaluation will be conducted using performance metrics such as accuracy, precision, and recall. It is expected, ultimately, to determine which of the models shows greater effectiveness for the proposed problem, contributing to the advancement of research in information security.

Keywords: Information Security. Phishing. Machine Learning. Data Classification.

---

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

Esta seção irá aprofundar os conceitos que servem de alicerce para o trabalho. Será abordada a natureza dos ataques de *phishing*, detalhando suas técnicas e características mais comuns. Em seguida, serão apresentados os fundamentos do Machine Learning, com foco em Aprendizado Supervisionado e no problema de Classificação, tendo como base a obra de **hastie2009elements**. Por fim, será realizada uma explanação conceitual sobre o funcionamento dos algoritmos selecionados e uma revisão de trabalhos correlatos, como o estudo de **mandadi2022**, que aplicaram técnicas similares para a detecção de phishing.

3 METODOLOGIA PRELIMINAR

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

1. **Fonte de Dados:** Será utilizado o dataset público “Phishing Dataset for Machine Learning” (**kaggle_dataset_2022**). 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 será 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:** Será realizada uma análise exploratória para compreender a distribuição e correlação dos dados. Posteriormente, o conjunto de dados será 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 para garantir a separabilidade e a avaliação imparcial do modelo.
4. **Modelagem e Avaliação:** Serão treinados e avaliados, a princípio, os algoritmos Random Forest e Support Vector Machine. A performance dos modelos será comparada utilizando as métricas de Acurácia, Matriz de Confusão, Precisão e Recall, a fim de determinar a abordagem mais eficaz para o problema proposto.

4 CRONOGRAMA

O desenvolvimento do projeto está planejado para ser executado ao longo de um semestre letivo, conforme a Tabela **??**.

Tabela 1 – Cronograma de Execução do Projeto.

Atividade	**Mês 1**	**Mês 2**	**Mês 3**	**Mês 4**	**Mês 5**
Revisão Bibliográfica	X	X			
Análise/Preparação dos Dados		X			
Desenvolvimento/Treinamento			X	X	
Análise dos Resultados/Escrita				X	X
Revisão Final e Entrega					X

5 RESULTADOS ESPERADOS

Espera-se que os modelos de Machine Learning treinados demonstrem alta capacidade preditiva na identificação de websites de phishing. Com base na literatura correlata, como em **mandadi2022**, projeta-se que ambos os algoritmos, Random Forest e SVM, alcançarão uma acurácia superior a 90%. O resultado da análise comparativa permitirá indicar qual modelo possui um desempenho mais robusto para este problema específico, considerando não apenas a acurácia geral, mas também a capacidade de minimizar falsos negativos, que representam o maior risco ao usuário.

6 CONSIDERAÇÕES FINAIS

Ao término deste trabalho, pretende-se entregar não apenas um estudo comparativo, mas também um modelo funcional de classificação que sirva como prova de conceito para ferramentas de segurança mais avançadas. A pesquisa visa contribuir para a comunidade acadêmica com uma análise metodológica clara e resultados reprodutíveis, além de reforçar a importância da aplicação de técnicas de Inteligência Artificial para a solução de problemas práticos e relevantes em segurança da informação.