INSTITUTO FEDERAL
Pernambuco

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 (Hastie; Tibshirani; Friedman, 2009). 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 (Mandadi *et al.*, 2022).

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

O Phishing constitui uma das modalidades de ataque cibernético mais difundidas e danosas da atualidade, sendo classificado como uma forma de engenharia social. O termo, um homófono da palavra inglesa fishing (pesca), alude de forma análoga à prática de ”pescar”informações confidenciais de usuários desatentos. Diferentemente de ataques que exploram vulnerabilidades técnicas em sistemas, o phishing visa o elo mais fraco da cadeia de segurança: o ser humano (Mandadi *et al.*, 2022). O objetivo principal de um ataque de phishing é enganar a vítima para que ela revele, voluntariamente, dados sensíveis como credenciais de acesso (nomes de usuário e senhas), números de cartão de crédito, informações bancárias e dados pessoais.

A eficácia do phishing reside na sua capacidade de explorar gatilhos psicológicos. Os perpetradores do ataque se passam por entidades confiáveis — como bancos, instituições governamentais, empresas de tecnologia ou até mesmo contatos conhecidos da vítima — para criar uma falsa sensação de legitimidade e segurança. As mensagens utilizadas no estratagema frequentemente apelam para um senso de urgência, medo ou curiosidade, compelindo a vítima a agir de forma impulsiva. Frases como ”Sua conta foi comprometida, clique aqui para verificar”, ”Você tem uma encomenda pendente”ou ”Sua assinatura expirou”são exemplos clássicos dessa tática.

Os ataques de phishing podem ser categorizados em diversas modalidades, que variam em seu nível de sofisticação e no seu alvo. As técnicas mais comuns são:

- **Phishing por E-mail:** É a forma mais tradicional e massificada. Os atacantes disparam milhões de e-mails genéricos para uma vasta lista de endereços, na esperança de que uma pequena porcentagem das vítimas clique no link malicioso. Esses e-mails costumam imitar a identidade visual de empresas conhecidas, mas geralmente contêm erros sutis de gramática ou de design.
- **Spear Phishing (Phishing Direcionado):** Uma evolução muito mais perigosa do ataque tradicional. No Spear Phishing, o atacante realiza uma pesquisa prévia sobre o alvo (uma pessoa ou um grupo específico dentro de uma organização) e personaliza a mensagem com informações relevantes para a vítima, como seu

nome, cargo, ou referências a projetos em que está trabalhando. Essa personalização aumenta drasticamente a credibilidade do ataque e, consequentemente, sua taxa de sucesso.

- **Whaling (Caça à Baleia):** Uma subcategoria de Spear Phishing que visa especificamente executivos de alto escalão (CEOs, CFOs, etc.), apelidados de ”peixes grandes”ou ”baleias”. Os ataques de Whaling são altamente personalizados e costumam ter como objetivo o roubo de informações estratégicas da empresa ou a autorização de transferências financeiras fraudulentas.
- **Smishing e Vishing:** São variações do phishing que utilizam, respectivamente, mensagens de texto (SMS) e chamadas de voz (VoIP). O Smishing frequentemente envolve links para aplicativos maliciosos ou sites falsos, enquanto o Vishing utiliza a interação humana para manipular a vítima e extrair informações por telefone.

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” (Tiwari, 2022), cuja base original foi apresentada por Mandadi *et al.* (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 foi realizado na linguagem Python, com o auxílio das bibliotecas Pandas para manipulação de dados e Scikit-learn (Pedregosa *et al.*, 2011) 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 os resultados quantitativos obtidos a partir da execução da metodologia, detalhando o desempenho individual de cada modelo de

Machine Learning. Em seguida, é realizada uma discussão aprofundada, comparando os modelos e interpretando os achados à luz do problema de pesquisa.

4.1 Desempenho dos Modelos

Após a etapa de treinamento com 80% dos dados, os modelos Random Forest e Support Vector Machine (SVM) foram avaliados com o conjunto de teste de 2000 amostras, que não haviam sido expostas aos modelos previamente. O desempenho de cada classificador foi mensurado utilizando métricas padrão de avaliação, cujos resultados estão consolidados na Tabela 1.

Tabela 1 – Resultados comparativos de desempenho dos modelos.

Métrica	Random Forest	SVM
Desempenho Geral		
Acurácia	**98.20%**	86.35%
Análise de Erros (do conjunto de teste com 2000 amostras)		
Phishing class. como Legítimo (FP)	**18**	181
Legítimo class. como Phishing (FN)	**18**	92
Desempenho na Classe "Phishing"		
Precisão	**0.98**	0.90
Recall	**0.98**	0.82
F1-Score	**0.98**	0.86

Fonte: O autor (2025)

O modelo Random Forest demonstrou um desempenho excepcional, alcançando uma acurácia geral de 98.20%. A análise de sua matriz de confusão revelou um equilíbrio notável, com apenas 18 erros de Falsos Positivos e 18 erros de Falsos Negativos.

O modelo SVM, por sua vez, obteve uma acurácia de 86.35%. Embora seja um resultado considerável, é significativamente inferior ao do Random Forest. Sua matriz de confusão indicou um número muito maior de erros, totalizando 181 Falsos Positivos e 92 Falsos Negativos.

4.2 Análise Comparativa e Discussão

A análise comparativa dos resultados apresentados na Tabela 1 indica uma clara superioridade do modelo Random Forest sobre o SVM para a tarefa de detecção de phishing neste conjunto de dados. A diferença de quase 12 pontos percentuais na acurácia é o primeiro indicador dessa disparidade.

O ponto mais crítico da comparação reside na análise dos erros, especificamente os Falsos Positivos — casos em que um site de phishing é incorretamente classificado como legítimo. O SVM cometeu 181 erros deste tipo, 10 vezes mais que os 18 erros do Random Forest. Em uma aplicação de segurança real, essa diferença é inaceitável, pois representa uma falha grave na proteção ao usuário.

A métrica de Recall para a classe "Phishing"reforça essa conclusão. O Random Forest obteve um Recall de 0.98, indicando que foi capaz de identificar 98% de todos os

sites maliciosos presentes no conjunto de teste. Em contrapartida, o SVM alcançou um Recall de 0.82, o que significa que 18% das ameaças não foram detectadas pelo modelo. A eficácia de uma ferramenta de segurança está diretamente ligada à sua capacidade de minimizar ameaças não detectadas, tornando o Random Forest a escolha mais confiável.

A superioridade do Random Forest pode ser atribuída à sua natureza de *ensemble*, que combina múltiplas árvores de decisão para criar um modelo mais robusto e menos propenso a superajuste (*overfitting*). Por outro lado, o desempenho inferior do SVM pode estar relacionado à sua sensibilidade a dados não escalonados e à necessidade de uma otimização de hiperparâmetros mais complexa, como documentado por Hastie, Tibshirani e Friedman (2009). O resultado de 98.20% obtido pelo Random Forest é, ainda, consistente com os achados de Mandadi *et al.* (2022), que também apontam o algoritmo como uma abordagem de alta performance para esta tarefa.

5 CONCLUSÃO

5.1 Síntese dos Resultados

5.2 Limitações do Trabalho

5.3 Trabalhos Futuros

REFERÊNCIAS

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. [*S. l.*]: Springer Science & Business Media, 2009.

MANDADI, Adarsh *et al.* Phishing Website Detection Using Machine Learning*. In:* 2022 IEEE 7th International conference for Convergence in Technology (I2CT). [*S. l.: s. n.*], 2022. p. 1–4. DOI: 10.1109/I2CT54291.2022.9824801.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

TIWARI, Shashwat. **Phishing Dataset for Machine Learning**. [*S. l.*]: Kaggle, 2022. Disponível em: https://www.kaggle.com/datasets/shashwatwork/phishing-dataset-for-machine-learning. Acesso em: 6 jul. 2025.