DETEKSI *WEBSITE PHISING* MENGGUNAKAN TEKNIK DATA *MINING* DAN *WEB CRAWLING*

Noviana Wulandari[1], Chanief Budi Setiawan[2]

INTISARI

Latar Belakang: *Phising* merupakan salah satu bentuk serangan siber yang semakin meningkat seiring dengan berkembangnya teknologi digital, dengan lebih dari 1,2 juta kasus global dan kerugian mencapai USD 42 miliar pada tahun 2023. Penelitian ini mengembangkan sistem deteksi situs *phising* menggunakan algoritma *Decision Tree* CART dan teknik *web crawling*, dengan data *real-time* dari *PhishTank* dan *Tranco*. CART dipilih karena kemampuannya menangani data numerik dan kategorikal secara efisien serta mudah diinterpretasi.

Tujuan: Penelitian ini bertujuan untuk melakukan evaluasi model deteksi *website phising* dengan menggunakna teknik data *mining* dan *web crawling* berbasis algoritma *Decision Tree CART* dengan memanfaatkan data *real-time*.

Metode Penelitian: Penelitian dilakukan melalui tahapan studi literatur, pengumpulan data dengan *web crawling*, *preprocessing* dan normalisasi, ekstraksi fitur, pelatihan model CART, evaluasi menggunakan *confusion metrix*.

Hasil: Model yang dibangun menghasilkan akurasi data uji sebesar 99,49%, data latih 99,57%, dan rata-rata akurasi *cross-validation* sebesar 99,56%. *Precision* dan *recall* untuk kedua kelas mencapai nilai 0,99, serta f1-*score* yang tinggi menunjukkan kestabilan dan keseimbangan model. Uji prediksi *real-time* terhadap 10 URL baru menunjukkan bahwa model mampu mengeneralisasi data secara akurat.

Kesimpulan: Model *Decision Tree* CART terbukti mampu mendeteksi situs *phising* dengan akurat dan efektif. Integrasi data *real-time* melalui *web crawling* dari *PhishTank* memungkinkan model tetap adaptif terhadap tren serangan terbaru, serta cocok diterapkan dalam sistem keamanan siber berbasis URL.

Kata-kunci: *Phising*, *Decision Tree*,CART, *Data Mining*, *Web Crawling*, Deteksi URL

---

[1]Mahasiswa Teknologi Informasi (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta
[2]Dosen Teknologi Informasi (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta

DETEKSI *WEBSITE PHISING* MENGGUNAKAN TEKNIK DATA *MINING* DAN *WEB CRAWLING*

Noviana Wulandari[1], Chanief Budi Setiawan[2]

ABSTRACT

***Background**: Phishing is a form of cyberattack that is increasing along with the development of digital technology, with more than 1.2 million global cases and losses reaching USD 42 billion in 2023. This research develops a phishing site detection system using the Decision Tree CART algorithm and web crawling techniques, with real-time data from PhishTank and Tranco. CART was chosen because of its ability to handle numerical and categorical data efficiently and easily interpreted.*

***Objective**: This study aims to evaluate the phishing website detection model using data mining and web crawling techniques based on the Decision Tree CART algorithm by utilizing real-time data.*

***Method**: The research was conducted through the stages of literature study, data collection with web crawling, preprocessing and normalization, feature extraction, CART model training, evaluation using confusion metrics.*

***Result**: The developed model achieved 99.49% accuracy on test data, 99.57% on training data, and an average cross-validation accuracy of 99.56%. Precision and recall for both classes reached 0.99, and a high f1-score indicated the model's stability and balance. Real-time prediction tests on 10 new URLs demonstrated the model's ability to accurately generalize the data.*

***Conclusion**: The CART Decision Tree model has been proven to detect phishing sites accurately and effectively. The integration of real-time data through PhishTank's web crawling allows the model to remain adaptive to the latest attack trends and is suitable for implementation in URL-based cybersecurity systems..*

***Keywords**: Phising, Decision Tree, CART, Data Mining, Web Crawling, URL Detection.*

---

[1]*Undergraduate Student of Information Technology, Faculty of Engineering and Information Technology, Universitas Jenderal Achmad Yani Yogyakarta*
[2]*Lecturer of Information Technology, Faculty of Engineering and Information Technology, Universitas Jenderal Achmad Yani Yogyakarta*