JURNAL ILMU KOMPUTER
Volume 3 Nomor 1, Juli, 2025
ISSN: 3031-125X (ONLINE)
Penerbit: Program Studi Teknik Informatika S-2 Universitas Pamulang

KLASIFIKASI *PHISHING* URL PADA *WEBSITE* BERBASIS METODE *ENSEMBLE*

Bahrul Ulum [1], Taswanda Taryo[2], dan Sudarno[3]
Program Studi Teknik Informatika S-2, Universitas Pamulang,
Email: ulumhm1@gmail.com[1], taswandataryo@gmail.com[2], dan sudarnowiharjo@gmail.com [3]

ABSTRACT

This study analyzes the performance of ensemble learning algorithms in detecting phishing URLs using the PhiUSIIL Phishing URL dataset. The three algorithms compared are CatBoost, XGBoost, and LightGBM. The research stages include data preprocessing, data division into an 80:20 train-test split, and performance evaluation based on accuracy, precision, recall, and F1-score metrics. The results show that XGBoost has the best performance with an accuracy of 97.54% and an ROC AUC of 93.05%, followed by CatBoost with an accuracy of 97.46% and an ROC AUC of 92.94%. LightGBM, although it has lower performance, still shows good results with an accuracy of 96.99% and an ROC AUC of 91.85%. The data cleaning process successfully improves efficiency by eliminating irrelevant attribute analysis. This study confirms that ensemble algorithms can be implemented for the development of more effective and accurate phishing detection systems. XGBoost is recommended as the primary algorithm in detecting phishing threats in cybersecurity applications, thanks to its ability to handle large and complex data.

Keywords: *phishing URL, ensemble learning, CatBoost, XGBoost, LightGBM, security detection, cybersecurity.*

ABSTRAK

Penelitian ini menganalisis kinerja algoritma ensemble learning dalam mendeteksi URL phishing menggunakan dataset PhiUSIIL Phishing URL. Tiga algoritma yang dibandingkan adalah CatBoost, XGBoost, dan LightGBM. Tahapan penelitian meliputi preprocessing data, pembagian data menjadi train-test split 80:20, dan evaluasi kinerja berdasarkan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa XGBoost memiliki performa terbaik dengan akurasi 97,54% dan ROC AUC 93,05%, diikuti oleh CatBoost dengan akurasi 97,46% dan ROC AUC 92,94%. LightGBM, meskipun memiliki performa lebih rendah, tetap menunjukkan hasil yang baik dengan akurasi 96,99% dan ROC AUC 91,85%. Proses cleansing data berhasil meningkatkan efisiensi analisis dengan menghilangkan atribut tidak relevan. Penelitian ini menegaskan bahwa algoritma ensemble dapat diimplementasikan untuk pengembangan sistem deteksi phishing yang lebih efektif dan akurat. XGBoost direkomendasikan sebagai algoritma utama dalam mendeteksi ancaman phishing pada aplikasi keamanan siber, berkat kemampuannya dalam menangani data besar dan kompleks.

Kata Kunci: phishing URL, ensemble learning, CatBoost, XGBoost, LightGBM, deteksi keamanan, keamanan siber.

1. PENDAHULUAN

Pandemi Covid-19 telah membawa perubahan besar pada kebiasaan dan gaya hidup masyarakat, termasuk peralihan aktivitas dari luring ke daring melalui berbagai platform digital. Internet memainkan peran sentral, tidak hanya sebagai alat pendukung aktivitas, tetapi juga sebagai sarana utama interaksi dan transaksi[1]. Evolusi website menjadi contoh nyata perubahan ini, dari sekadar media penyampaian informasi satu arah menjadi platform interaktif dan transaksional. Namun, perkembangan ini juga memunculkan ancaman keamanan, salah satunya adalah serangan web phishing, yang menjadi perhatian global karena dampaknya yang merugikan pengguna[2]. .

Laporan Anti-Phishing Working Group (APWG) pada kuartal ketiga 2021 mencatat lebih dari 700.000 situs phishing aktif dan jutaan klik pada tautan phishing[3]. Untuk menghadapi ancaman ini, diperlukan sistem pendeteksian yang andal. Penelitian di bidang data mining dan machine learning terus dikembangkan untuk mendeteksi serangan phishing[4]. Algoritma seperti Decision Tree, Random Forest, dan Support Vector Machine telah menunjukkan tingkat akurasi yang tinggi dalam deteksi phishing, dengan Random Forest mencapai akurasi 97,14% pada beberapa studi[5].

Meskipun demikian, masih jarang penelitian yang mengembangkan algoritma machine learning menjadi aplikasi deteksi phishing berbasis web[6]. Penelitian terkait algoritma Decision Tree menunjukkan hasil yang stabil pada dataset tertentu, sementara metode pengembangan perangkat lunak seperti Waterfall terbukti efektif dalam implementasi sistem serupa. Penelitian yang berfokus pada metode ensemble, seperti CatBoost, XGBoost, dan LightGBM, menawarkan potensi besar karena kemampuan algoritma ini dalam menangani data yang besar dan kompleks[7].

Penelitian ini bertujuan untuk menganalisis kinerja metode ensemble dalam mendeteksi URL phishing, menggunakan dataset PhiUSIIL Phishing URL. Metode yang dipilih akan dievaluasi berdasarkan akurasi, presisi, recall, dan F1-score untuk menentukan algoritma paling efektif. Hasil penelitian diharapkan memberikan rekomendasi penggunaan metode ensemble dalam keamanan siber, khususnya pengembangan sistem deteksi phishing yang lebih efisien dan akurat di masa depan[8].

2. METODE

Penelitian ini diawali dengan analisis kebutuhan, yang mencakup perangkat keras, perangkat lunak, dan data. Perangkat keras yang digunakan adalah komputer dengan spesifikasi minimal prosesor Intel Core i5-7300U, RAM 16 GB, dan harddisk SSD 240 GB[9]. Perangkat lunaknya meliputi sistem operasi Windows 10 Pro (versi 22H2), aplikasi Google Colab, dan bahasa pemrograman Python. Dataset yang digunakan adalah PhiUSIIL Phishing URL Dataset dari UCI Machine Learning Repository. Tahapan penelitian meliputi identifikasi masalah, kajian pustaka untuk memahami metode deteksi phishing sebelumnya, pengumpulan data, preprocessing berupa pembersihan dan transformasi data, hingga pembagian dataset menjadi 80% data latih dan 20% data uji[10].

Selanjutnya, penelitian memanfaatkan teknik ensemble learning menggunakan algoritma CatBoost, XGBoost, dan LightGBM untuk mengolah data. Evaluasi model dilakukan melalui berbagai metrik, seperti akurasi, presisi, recall, dan F1-score. Data yang telah diproses diuji menggunakan metode train-test split untuk menilai performa model secara terpisah. Langkah ini bertujuan untuk memperoleh model yang andal dalam mendeteksi URL phishing dan memberikan rekomendasi terkait algoritma yang paling efektif dalam meningkatkan keamanan siber.

3. HASIL

3.1. Data Preprocessing

Dataset Phishing URL yang diunduh dari UCI Machine Learning Repository awalnya memiliki 57 kolom dan 235.795 baris data. Setelah dilakukan proses pembersihan (cleansing), dataset ini menyusut menjadi 52 kolom dan 199.159 baris. Dataset ini dirancang untuk mendeteksi dan mengklasifikasikan URL phishing berdasarkan berbagai fitur, dengan beberapa fitur utama seperti Label (kategori phishing atau non-phishing untuk klasifikasi), Attack Type (jenis ancaman spesifik seperti phishing atau malware), serta tipe data int64 untuk atribut numerik dan Object untuk informasi teks seperti nama domain atau URL. Dalam proses cleansing, beberapa kolom dihapus karena dianggap tidak relevan atau menambah kompleksitas tanpa kontribusi signifikan, seperti kolom FILENAME (metadata nama file), URL (URL lengkap yang sulit diproses dan berisiko overfitting), Domain (informasi domain yang tercakup dalam

fitur lain), TLD (informasi umum domain), dan Title (judul halaman yang membutuhkan pemrosesan tambahan seperti NLP).

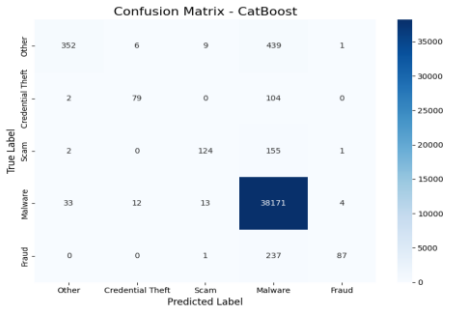
Dataset ini juga mencakup fitur-fitur yang berfokus pada struktur URL (misalnya panjang URL, jumlah subdomain, dan rasio huruf atau angka), keamanan dan komponen halaman (seperti penggunaan HTTPS, keberadaan favicon, dan elemen form), serta karakteristik teknikal (jumlah baris kode, file CSS, JS, dan referensi eksternal). Cleansing dataset dilakukan untuk menghapus kolom yang tidak informatif, redundan, atau berupa data non-fitur, seperti metadata. Dengan fokus pada fitur-fitur yang relevan, dataset ini menjadi sumber penting dalam analisis ancaman phishing dan pengembangan model pembelajaran mesin.

3.2. Hasil

Setelah dilakukan pembersihan data maka dapat dipastikan dataset yang digunakan saat ini, untuk semua kolom tidak memiliki nilai kosong (non-null), menunjukkan bahwa proses pembersihan telah dilakukan dengan baik. Langkah selanjutnya melakukan pelatihan dan pengujian dari masing-masing model ensemble (CatBoost, XGBoost, dan LightGBM).

a. CatBoost *(Categorical Boosting)*

Berikut adalah hasil pengujian dari masing-masing model yang telah dianalisis, yang pertama dari model *CatBoost (Categorical Boosting)* lihat pada Gambar 1 di bawah ini:



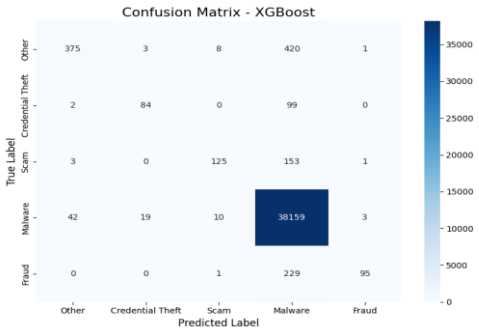
Gambar 1. Confusion Matrix – CatBoost

Confusion Matrix – CatBoost (Categorical Boosting) diatas menunjukkan performa model dalam mengklasifikasikan data ke dalam lima kategori: *"Other, Credential Theft, Scam, Malware, dan Fraud"*. Untuk elemen-elemen dalam gambar tersebut dapat dijelaskan sebagai berikut:

- Sumbu Vertikal (*True Label*): Merupakan kategori data sebenarnya dari yang diuji.
- Sumbu *Horizontal* (*Predicted Label*): Merupakan model kategori yang diprediksi.
- Nilai Sel *Matriks*: Setiap sel merupakan jumlah sample yang sesuai dengan kombinasi label sebenarnya (sumbu *vertical*) dan prediksi model (sumbu *horizontal*).

b. Confusion Matrix – XGBoost (Extreme Gradient Boosting)

Selanjutnya hasil kedua dari model *Confusion Matrix – XGBoost (Extreme Gradient Boosting)*, dapat dilihat pada Gambar 2:



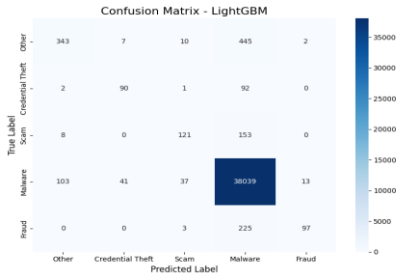
Gambar 2. Confusion Matrix – XGBoost

Confusion Matrix – XGBoost (Extreme Gradient Boosting) diatas menunjukkan performa model dalam mengklasifikasikan data ke dalam lima kategori: "*Other, Credential Theft, Scam, Malware,* dan *Fraud*". Untuk elemen-elemen dalam gambar tersebut dapat dijelaskan sebagai berikut:

- Sumbu Vertikal (*True Label*): Merupakan kategori data sebenarnya dari yang diuji.
- Sumbu Horizontal (*Predicted Label*): Merupakan model kategori yang diprediksi.
- Nilai Sel Matriks: Setiap sel merupakan jumlah sample yang sesuai dengan kombinasi label sebenarnya (sumbu *vertical*) dan prediksi model (sumbu *horizontal*).

c. *Confusion Matrix – LightGBM (Light Gradient Boosting Machine)*

Selanjutnya hasil ketiga dari model *Confusion Matrix – LightGBM (Light Gradient Boosting Machine)*, dapat dilihat pada Gambar 3:



Gambar 3. Confusion Matrix - LightGBM

Confusion Matrix – LightGBM (Light Gradient Boosting Machine) diatas menunjukkan performa model dalam mengklasifikasikan data ke dalam lima kategori: "*Other, Credential Theft, Scam, Malware*, dan *Fraud*". Untuk elemen-elemen dalam gambar tersebut dapat dijelaskan sebagai berikut:

- Sumbu Vertikal (*True Label*): Merupakan kategori data sebenarnya dari yang diuji.
- Sumbu Horizontal (*Predicted Label*): Merupakan model kategori yang diprediksi.
- Nilai Sel Matriks: Setiap sel merupakan jumlah sample yang sesuai dengan kombinasi label sebenarnya (sumbu *vertical*) dan prediksi model (sumbu *horizontal*).

Hasil dari ketiga model yang telah dianalisis dapat di lihat perbandingan prediksi secara mendetail dari tingkat nilai akurasi dan Tingkat nilai ROC AUC pada tabel 1:

Tabel 1. Klasifikasi Kinerja Boosting

Classifier	**Accuracy**	**Persentase**	**ROC AUC**	**Persentase**
CatBoost	0.974618	97,46%	0.929431	92,94%
XGBoost	0.975422	97,54%	0.930519	93,05%
LightGBM	0.969949	96,99%	0.918547	91,85%

Dari tabel 1 tersebut menunjukan perbndingan kinerja dari tiga classifier yang berbeda (Catboost, XGBoost, dan LightGBM) dengan pembagian data *train-test split* 80:20. Tabel tersebut mengukur kinerja dari setiap kombinasi preprocessing dan classifier berdasarkan akurasi (Acc) dan ROC AUC. Berikut pejelasan detail dari tabel tersebut:

1. **Classifier:** Nama algoritma yang digunakan untuk melakukan klasifikasi data.
2. **Accuracy:** Tingkat ketepatan model, dihitung sebagai rasio prediksi benar terhadap total jumlah prediksi.

   Disajikan dalam bentuk desimal (misalnya, 0.974618) dan persentase (misalnya, 97,46%).

3. **ROC AUC:** Area Under the Curve pada Receiver Operating Characteristic, mengukur kemampuan model untuk membedakan antara kelas positif dan negatif. Disajikan dalam bentuk desimal (misalnya, 0.929431) dan persentase (misalnya, 92,94%).

Penjelasan Detail Tabel:

1. CatBoost:
   a. TTS 80:20: Akurasi: 0.974618 (97,46%).
   b. TTS 80:20: ROC AUC: 0.929431 (92,94%).

   Keterangan:

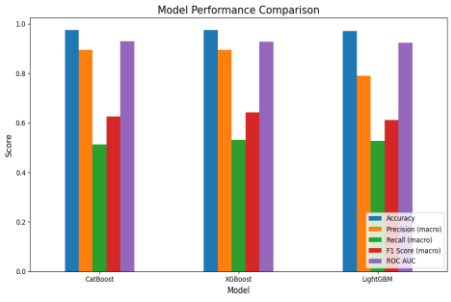
   Model ini memprediksi dengan akurasi tinggi (97,46%), dan nilai ROC AUC sebesar 92,94% menunjukkan kemampuan yang baik dalam membedakan antara kelas.
2. XGBoost:
   a. TTS 80:20: Akurasi: 0.975422 (97,54%).
   b. TTS 80:20: ROC AUC: 0.930519 (93,05%).

   Keterangan:

   XGBoost memiliki nilai akurasi tertinggi (97,54%), dan nilai ROC AUC (93,05%). XGBoost lebih tinggi dari CatBoost.
3. LightGBM:
   a. TTS 80:20: Akurasi: 0.969949 (96,99%).
   b. TTS 80:20: ROC AUC: 0.918547 (91,85%).

   Keterangan:

   Model ini memiliki akurasi paling rendah (96,99%) dibandingkan dua model lainnya. ROC AUC juga sedikit lebih rendah (91,85%), namun tetap berada pada kisaran yang baik.

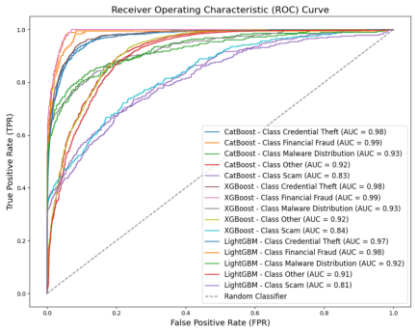


Gambar 4. Model Performance Comparison

Dari hasil ketiga model essemble diatas, berikut adalah evaluasi dan perbandingan kemampuan model machine learning atau algoritma untuk mengetahui model terbaik dapat dilhat pada Gambar 7. Dari gambar diatas, setiap model machine learning memiliki tingkat performa yang hampir sama. Namun tetap dapat dibedakan model terbaik dari *(CatBoost, XGBoost, LightGBM).*

1. Accuracy - (Bar Biru)
   a) Ketiga model memiliki akurasi yang tinggi.
   b) Akurasi XGBoost berada sedikit lebih tinggi dibandingkan CatBoost dan LightGBM.
2. Precision (macro) - Bar Oranye
   a) LightGBM memiliki precision lebih rendah dibandingkan dengan XGBoost dan CatBoost.
3. Recall (macro) – Bar Hijau
   a) Dibandingkan metrik lainnya recall (macro) memiliki nilai yang lebih rendah, terutama pada model LightBM.
   b) Kinerja LightGBM pada metrik ini lebih buruk dibandingkan kinerja dari XGBoost dan CatBoost.
4. F1 Score (macro) - Bar Merah
   a) F1 Score mendekati antara precision dan recall, dengan pola yang serupa.
   b) Dibandingkan LightGBM, lebih unggul XGBoost dan CatBoost.
5. ROC AUC - Bar Ungu
   a) Untuk semua model mendapatkan skor ROC AUC yang tinggi.
   b) Terendah LightGBM, CatBoost lebih tinggi dari LightGBM, dan XGBoost memiliki skor tertinggi dari keduanya.

Receiver Operating Characteristic (ROC) Curve menunjukan performa dari pembelajaran mesin dalam hal ini mengacu pada tiga model (CatBoost, XGBoost, LightGBM) untuk mengklasifikasikan beberapa kelas yaitu *Other, Credential Theft, Scam, Malware, Fraud.* Lihat pada Gambar 4.7:



Gambar 5. Receiver Operating Characteristic (ROC) Curve

Gambar diatas dapat dijelaskan sebagai berikut:

1. Sumbu X (False Positive Rate - FPR): Menunjukan tingkat kesalahan positif (Proporsi sampel negatif yang salah diklasifikasikan sebagai positif).
2. Sumbu Y (True Positive Rate - TPR): Menunjukan tingkat keberhasilan positif (Proporsi sampel positif yang benar diklasifikasikan sebagai positif).
3. Kurva ROC: Mewakili performa model pada setiap kelas. Semakin mendekati pojok kiri atas, semakin baik performa model karena itu menunjukan TPR yang tinggi dan FPR yang rendah.
4. AUC (Area Under the Curve): Nilai yang menunjukan seberapa baik model dalam memisahkan kelas. Nilai AUC berkisar antara 0 dan 1, dengan:
   - AUC mendekati 1: Model sangat baik dalam memisahkan kelas.
   - AUC mendekati 0.5: Model tidak lebih baik dari pada tebakan acak.
5. Legenda: Menyediakan informasi tentang model, kelas, dan nilai AUC. Untuk model CatBoost, XGBoost, dan LightGBM dibandingkan untuk setiap kelas, dan semua model memiliki kinerja sangat baik (dengan AUC antara 0.81 hingga 0.99) berbeda dengan kelas “scam” karena memiliki performa lebih rendah.

Dashed Line (Random Classifier): Garis diagonal menunjuka performa pengklasifikasi acak dengan AUC=0.5 sebagai referensi.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode ensemble learning yang melibatkan algoritma CatBoost, XGBoost, dan LightGBM memiliki efektivitas tinggi dalam mendeteksi URL phishing. Dataset PhiUSIIL Phishing URL digunakan untuk melatih dan menguji model, dengan hasil akurasi dan nilai ROC AUC yang sangat baik untuk

semua algoritma. XGBoost mencapai performa terbaik dengan akurasi 97,54% dan ROC AUC 93,05%, diikuti oleh CatBoost dengan akurasi 97,46% dan ROC AUC 92,94%. LightGBM memiliki akurasi terendah di antara ketiganya, yaitu 96,99%, dan ROC AUC sebesar 91,85%, namun tetap menunjukkan performa yang baik.

Proses cleansing dataset berhasil menghilangkan atribut yang tidak relevan sehingga meningkatkan efisiensi analisis. Dataset ini berfokus pada fitur-fitur yang mendukung deteksi phishing, seperti struktur URL, komponen halaman, dan karakteristik teknikal lainnya. Hasil penelitian ini menegaskan bahwa algoritma ensemble dapat digunakan untuk mengembangkan sistem deteksi phishing yang lebih akurat dan andal. XGBoost direkomendasikan sebagai algoritma paling unggul dalam penelitian ini, menjadikannya pilihan utama untuk aplikasi keamanan siber dalam mendeteksi ancaman phishing.

5. REFERENSI

[1] M. Samantri and Afiyati, “Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 1–9, 2024, doi: 10.35870/jtik.v8i1.1202.

[2] I. A. Hidayat, “Classification of Sleep Disorders Using Random Forest on Sleep Health and Lifestyle Dataset,” *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 3, no. 2, pp. 71–76, 2023, doi: 10.20895/dinda.v3i2.1215.

[3] A. F. Mahmud and S. Wirawan, “Sistemasi: Jurnal Sistem Informasi Deteksi Phishing Website menggunakan Machine Learning Metode Klasifikasi Phishing Website Detection using Machine Learning Classification Method,” vol. 13, no. 4, pp. 2540–9719, 2024, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[4] A. Ferdita Nugraha, R. F. A. Aziza, and Y. Pristyanto, “Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing,” *J. Infomedia*, vol. 7, no. 1, p. 39, 2022, doi: 10.30811/jim.v7i1.2959.

[5] C. Umam and L. B. Handoko, “Prediksi Email Phising Menggunakan Support Vector Machine,” *Semnas Ristek (Seminar Nas. Ris. dan Inov. Teknol.*, vol. 8, no. 01, pp. 85–89, 2024, doi: 10.30998/semnasristek.v8i01.7138.

[6] A. D. Harahap, D. Juardi, and A. S. Y. Irawan, "Rancang Bangun Sistem Pendeteksi Link Phishing Menggunakan Algoritma Random Forest Berbasis Web," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4858.

[7] F. S. S. Nagalay, F. I. Komputer, S. Informasi, K. Siber, and W. Phishing, "ANALISIS PENERAPAN ALGORITMA DECISION TREE DALAM KEAMANAN SIBER UNTUK KELASIFIKASI SITUS WEBSITE," vol. 10, no. 1, pp. 1–8, 2024.

[8] T. T. F. Manguma and E. Fatra, "Analisis Performa Algoritma Klasifikasi untuk Deteksi Spam pada Email," *Innov. J. Soc. Sci. …*, vol. 4, pp. 16461–16465, 2024, [Online]. Available: http://j-innovative.org/index.php/Innovative/article/download/12547/8461

[9] C. M. Bachri and W. Gunawan, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Deteksi Email Spam menggunakan Algoritma Convolutional Neural Network (CNN)," *Edukasi dan Penelit. Inform.*, vol. 10, no. 1, pp. 88–94, 2024.

[10] P. Subarkah and A. N. Ikhsan, "Identifikasi Website Phishing Menggunakan Algoritma Classification And Regression Trees (CART)," *J. Ilm. Inform.*, vol. 6, no. 2, pp. 127–136, 2021, doi: 10.35316/jimi.v6i2.1342.