Journal of Computer Science and Technology
Volume 4 ; Nomor 2 ; Mei 2026 ; Page 52-59
DOI : https://doi.org/10.59435/jocstec.v4i2.669
https://jurnal.padangtekno.com/index.php/jocstec
E-ISSN : 2985-4318
P-ISSN : 2985-5772

Implementasi Algoritma Random Forest Untuk Deteksi Phishing Pada Email

Elfina Maulid[1*], Eko Amri Jaya[2], Muhamad Pradana Hardi Wasesa[3]

[1]Ilmu Komputer, Universitas Media Nusantara Citra
[2]Sistem Informasi, Universitas Media Nusantara Citra
[3]Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Indo Daya Suvana
[1*]dhearan.lili01@email.com

Abstrak

Serangan phishing melalui email merupakan ancaman cybersecurity yang terus berkembang dan dapat menyebabkan kerugian finansial maupun kebocoran data yang serius. Untuk mengatasi permasalahan ini, dibutuhkan metode yang andal dan otomatis guna membedakan email phishing dari email yang sah. Algoritma Random Forest dipilih karena kemampuannya dalam menangani struktur data yang kompleks dan menghasilkan akurasi klasifikasi yang tinggi. Pada penelitian ini, algoritma Random Forest diimplementasikan untuk klasifikasi email phishing menggunakan dataset yang terdiri dari email phishing dan email yang sah. Dataset tersebut diproses melalui tahapan pra-pemrosesan, ekstraksi fitur, pelatihan model, dan evaluasi. Hasil penelitian menunjukkan bahwa sistem berhasil mengklasifikasikan email phishing dan email yang sah dengan tingkat keakuratan yang tinggi, sehingga efektif dalam mendukung proses penyaringan email secara otomatis dan aman.

Kata Kunci : Phishing, Email, Random Forest, Klasifikasi, Cybersecurity

Abstract

Phishing attacks through email represent a growing cybersecurity threat that can lead to severe financial and data losses. To address this challenge, a reliable and automated method is required to distinguish phishing emails from legitimate ones. The Random Forest algorithm is selected due to its ability to handle complex data structures and deliver high classification accuracy. In this study, the Random Forest algorithm was implemented for phishing email classification using a dataset consisting of phishing and legitimate emails. The dataset was processed through stages of preprocessing, feature extraction, model training, and evaluation. The results indicate that the system successfully classified phishing and legitimate emails with a high level of accuracy, demonstrating its effectiveness in supporting automated and secure email filtering.

Keyword : *Phishing, Email, Random Forest, Classification, Cybersecurity*

PENDAHULUAN

Email merupakan salah satu media komunikasi utama dalam aktivitas pribadi, bisnis, hingga pemerintahan. Kemudahan dan kecepatan yang ditawarkan juga dimanfaatkan oleh pelaku kejahatan siber untuk melancarkan serangan phishing. Serangan ini merupakan bentuk penipuan digital yang memanfaatkan teknik rekayasa sosial (social engineering) untuk mengecoh korban agar menyerahkan data sensitif seperti kredensial akun maupun informasi keuangan[1]. Menurut laporan Phishing Activity Trends Report yang diterbitkan oleh Anti-Phishing Working Group (APWG), lebih dari 1,2 juta insiden phishing tercatat secara global pada kuartal ketiga tahun 2023, meningkat 38% dibanding periode sebelumnya. Sebagian besar serangan tersebut disebarkan melalui email,

mengingat jangkauannya yang luas serta kemudahannya untuk dimodifikasi secara masif [2]. Laporan lain dari FBI melalui Internet Crime Complaint Center (IC3) menunjukkan bahwa total kerugian akibat phishing pada tahun 2022 mencapai 3,4 miliar dolar Amerika Serikat, dengan sektor finansial menjadi salah satu target utama[3].

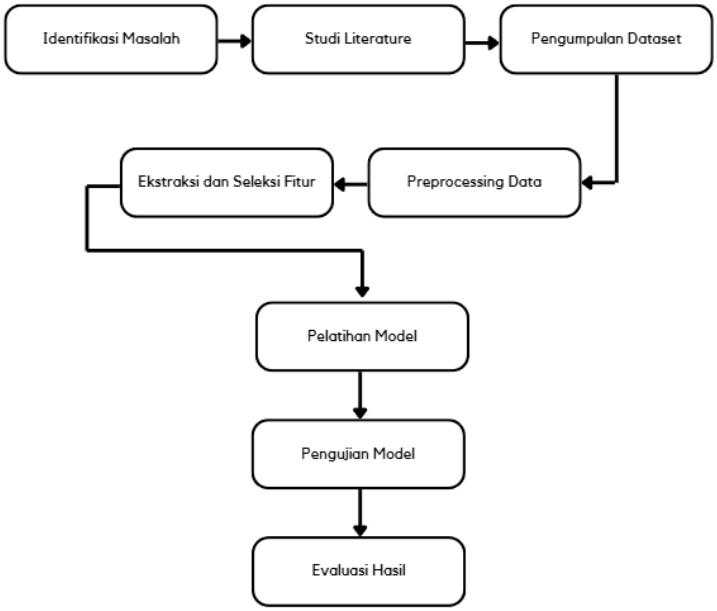
Secara otomatis pendekatan Machine Learning dapat digunakan melakukan deteksi phising pada email. Random Forest menjadi salah satu yang paling andal karena kemampuannya dalam menghasilkan prediksi yang akurat, stabil, dan tahan terhadap overfitting [4][5]. Penelitian [6] menunjukan Algoritma Random Forest mampu memberikan akurasi 98,10% dalam proses klasifikasi. Dalam deteksi phising pada website penelitian [7] menunjukan Algoritma Random Forest memiliki akurasi lebih baik dibanding algoritma lainya yaitu sebesar 99,5% , penelitian [8] sebesar 94,79% dan penelitian [9] sebesar 95,3%.

Penelitian ini bertujuan untuk mengimplementasi Algoritma Random Forest dalam mendeteksi phising pada email. Pemilihan algoritma Random Forest didasarkan atas pembuktian oleh peneliti sebelumnya. Data yang digunakan diperoleh dari PhishTank, SpamAssassin, dan Kaggle yang akan dilakukan proses normalisasi, parsing, kemudian dilanjutkan proses TF-IDF dan terakhir dilakukan proses klasifikasi menggunakan algortima Random Forest. Pola klasifikasi akan diimplementasi ke dalam aplikasi berbasis web agar dapat mendeteksi sebuah email terdeteksi phising atau nonphising.

METODOLOGI PENELITIAN

Tahapan Penelitian

Penelitian ini menerapkan pendekatan sistematis untuk menganalisis email phishing menggunakan algoritma Random Forest. Proses penelitian dimulai dari pengumpulan dataset hingga evaluasi kinerja model. Gambar. 1 berikut menggambarkan alur penelitian secara umum:



Gambar. 1 Alur Penelitian

Dari Gambar. 1 terlihat tahapan penelitian secara keseluruhan. Tahapan penelitian dimulai dengan melakukan identifikasi masalah, studi literature untuk menemukan penelitian terdahulu terkait topik, pengumpulan data, preprocessing data untuk melihat apakah data sudah bersih dan siap dilanjutkan ke tahap ekstraksi dan seleksi fitur. Selanjutkan dilakukan Pelatihan Model, pengujian model menggunakan Algoritma Random Forest. Tahapan terakhir adalah Evaluasi hasil untuk mengetahui apakah sistem dapat melakukan deteksi phising email secara akurat.

Alur penelitian yang terdapat pada Gambar. 1 akan dijelaskan secara rinci dengan memaparkan detail kegiatan, metode yang digunakan dan luaran pada setiap alur yang dapat dilihat pada Tabel. 1 berikut:

Tabel.1 Tahapan Penelitian

Tahap	Kegiatan Penelitian	Metode yang Digunakan	Luaran
1	Identifikasi Masalah	Identifikasi Masalah	Masalah yang akan dijadikan penelitian
2	Studi Literature	Studi Literature Review	Penelitian-penelitian terdahulu terkait algoritma dan metode
3	Pengumpulan Dataset	Dokumentasi/Data Sekunder	Dataset
4	Pre-processing Data	Cleaning, Parsing dan Split Data	Dataset
5	Ekstraksi dan Seleksi Fitur	TF-IDF	Dataset Numerik
6	Pelatihan Model	Algoritma Random Forest	Formula/Model Algoritma
7	Pengujian Model	Confusion Matrix	Presisi, Recall, dan F1-Score
8	Evaluasi Hasil	Anaslisi Deskriptif	Kesimpulan Penelitian

Tahapan identifikasi masalah dilakukan dengan menemukan masalah-masalah yang terjadi dalam phising email, disusun rumusan masalah dan penentuan tujuan penelitian. Selanjutnya pada tahapan Studi Literature dilakukan dengan mencari kajian-kajian terkait objek dan topik penenelitan melalui jurnal dan buku. Tahapan selanjutnya pengumpulan dataset yang diperoleh dari Keaggle, Phis Tank dan Spam Assasin yang merupakan data utama yang akan diolah. Tahapan selanjutnya dilakukan pre-processing data untuk mengetahui apakah data yang dikumpulkan sudah siap diolah. Pada tahapan ini dilakukan cleaning data untuk mencari data-data yang kosong, parsing data untuk memisahkan bagian header, body dan isi email [10]. Tahapan selanjutnya adalah Ekstraksi fitur menggunakan metode TF-IDF untuk mengubah data menjadi bentuk numeric [11]. Selanjutnya melakukan pelatihan model menggunakan algoritma Random Forest. Hasil dari pelatihan ini akan diimplementasikan ke dalam website sederhana, dengan memanggil hasil pelatihan. Tahapan selanjutnya pengujian mode dengan Confusion Matrik untuk melihat nilai presisi, Recall dan F1 Score dari algoritma Random Forest. Tahapan terakhir adalah Evaluasi Hasil untuk menarik kesimpulan dari penelitian.

Algoritma Random Forest

Algoritma machine learning Random Forest digunakan untuk klasifikasi serangan phishing, dengan tujuan utama mengembangkan pengklasifikasi email phishing yang lebih baik, memiliki akurasi prediksi yang lebih tinggi, serta menggunakan jumlah fitur yang lebih sedikit [12]. Algoritma Random Forest dapat dikembangkan dengan lebih mudah karena dukungan Scikit-learn yang menyediakan antarmuka API yang konsisten serta dokumentasi yang lengkap[13].

Untuk mendukung pemahaman terhadap mekanisme kerja algoritma ini, Random Forest dapat dirumuskan sebagai berikut. Misalkan terdapat T buah decision tree yang masing-masing dilambangkan sebagai $h_1(x)$, $h_2(x)$, ..., $hT(x)$, maka hasil akhir dari Random Forest adalah hasil voting mayoritas dari seluruh pohon:

Rumus: $H(x) = \arg\max\ y \in y \sum_{t=1}^{T} 1[h_t(x) = y]$ (1)

Selain itu, untuk menentukan atribut terbaik pada setiap node, digunakan indeks Gini. Rumus Gini dapat dituliskan sebagai: $\text{Gini(S)} = 1 - \sum_{i=1}^{c} pi^2$ (2)

Supervised Learning

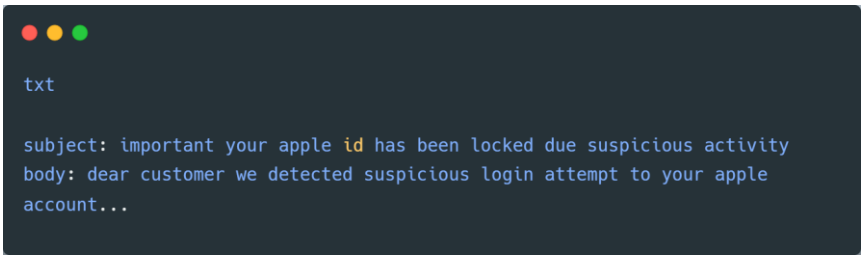
Supervised Learning merupakan metode machine learning yang memanfaatkan dataset berlabel untuk melatih model AI dalam mengenali pola serta hubungan yang mendasarinya. Tujuan utamanya adalah menghasilkan model yang mampu memprediksi hasil pada data baru di dunia nyata secara akurat [14].

Phising

Phishing secara tunggal merupakan tantangan, mengingat sifatnya yang terus berkembang dan beragamnya taktik yang digunakan oleh pelaku. Berbagai penelitian telah mencoba merumuskan definisi yang paling representatif. Salah satu upaya signifikan dalam mencapai konsensus mendefinisikan Phishing sebagai "tindakan penipuan yang dapat diskalakan di mana peniruan identitas digunakan untuk mendapatkan informasi dari target"[15].

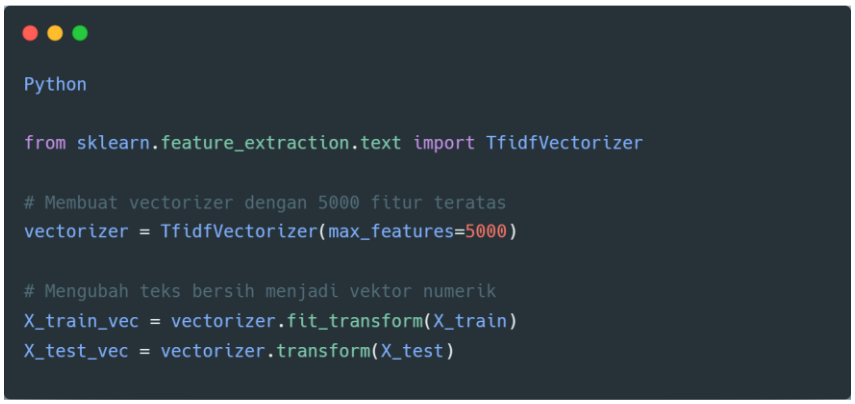
HASIL DAN PEMBAHASAN

Berdasarkan data yang telah diperoleh pada tahapan pengumpulan data, akan dilakukan proses presprocessing data dengan melakukan parsing dan split data. Hasil parsing data dapat dilihat pada gambar berikut:



Gambar.2 Hasil Parsing Data Email

Hasil parsing akan dilanjutkan dengan pembagian data menjadi 80% data training dan 20% data testing. Selanjut dilakukan Ektraksi dan Tranform Fitur dengan perintah sebagai berikut:

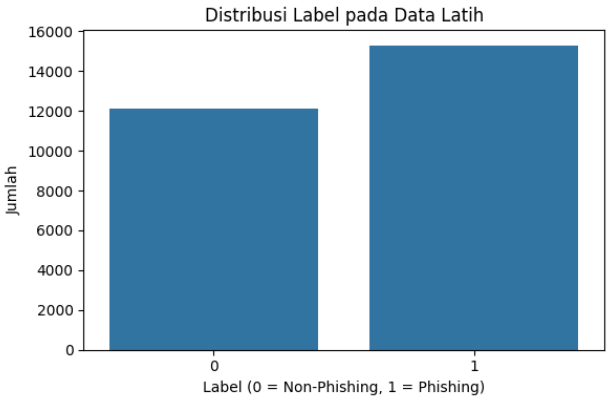


Gambar.3 Proses Ekstraksi dan Tranform Fitur

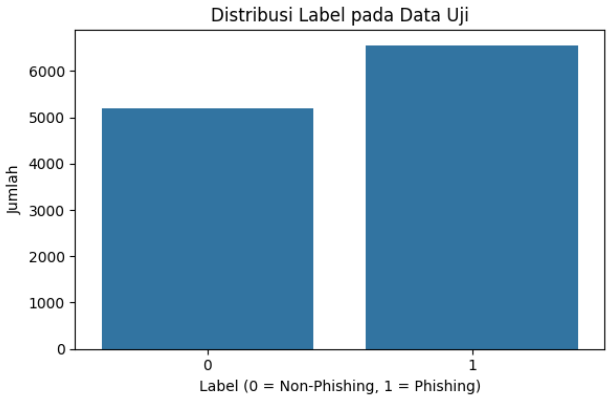
Hasil ekstraksi akan disimpan dalam bentuk code penyimpanan vectorizer dengan hasil **joblib.dumb(vectorizer, "tfidf_vectorizer.joblib").** Dengan adanya penyimpanan objek vectorizer, sistem dapat menjamin bahwa proses ekstraksi fitur dari data uji dilakukan dengan standar yang sama seperti pada data pelatihan. Hal ini sangat krusial karena jika terjadi perbedaan dalam representasi fitur, maka model tidak akan mampu melakukan prediksi dengan benar, bahkan bisa menyebabkan galat (error) saat proses inferensi.

Selanjutnya dilakukan pelatihan model menggunakan algoritma Random Forest. Dengan konfigurasi ini, proses pembagian data berhasil menjaga distribusi label di antara data latih dan data uji tetap seimbang, serta memastikan bahwa hasil evaluasi model dapat merepresentasikan kondisi

nyata. Visualisasi dari distribusi label pada masing-masing subset ditampilkan pada Gambar.5 dan Gambar.6 berikut:

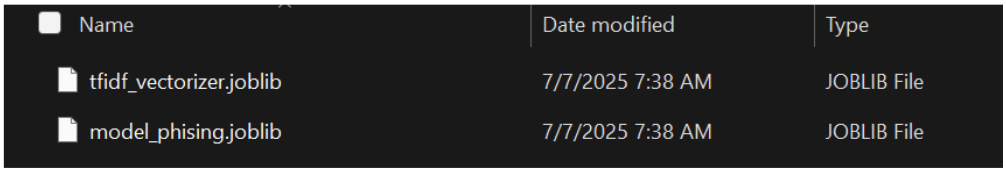


Gambar. 5 Distribusi Label pada Data Latih



Gambar. 6 Distribusi Label pada Data Uji

Pada pelatihan model Data Training menggunakan Algoritm Random Forest, penyimpanan model dilakukan dengan menggunakan pustaka joblib, yang merupakan standar umum untuk menyimpan objek Python berukuran besar seperti model Machine learning dan hasil vektorisasi. Hasil penyimpanan dapat dilihat pada gambar berikut:



Gambar. 7 Hasil Penyimpanan Model Pelatihan

Selanjutnya dilakukan pengujian menggunakan data testing, dan diperoleh hasil sebagai berikut:

Tabel. 2 Hasil Pengujian Terhadap Data Training Kaggle

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5194
1	0.99	0.99	0.99	6553
accuracy			0.99	11747
macro avg	0,99	0,99	0,99	11747
weighted avg	0.99	0,99	0,99	11747

Kelas Non-Phishing (label 0):

1. Precision: 1.00 – Menunjukkan bahwa semua prediksi email non-phishing adalah benar.
2. Recall: 1.00 – Semua email non-phishing yang sebenarnya berhasil dikenali oleh model.
3. F1-score: 1.00 – Kombinasi dari precision dan Recall yang menunjukkan kinerja sangat baik dalam mengenali email non-phising

Kelas Phishing (label 1):

1. Precision: 0,99 – Menunjukan 99% email yang diprediksi sebagai phishing memang benar-benar phishing.

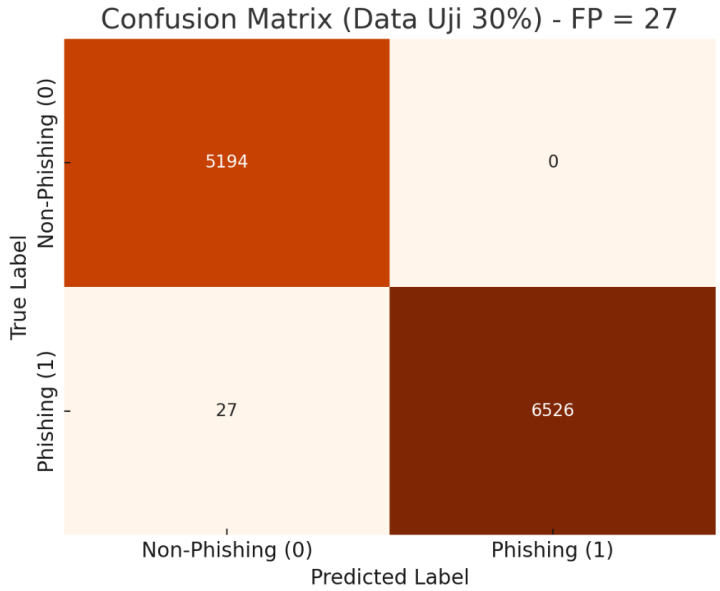
2. Recall: 0.99 – Menunjukkan 99% email phishing dikenali oleh model.
3. F1-Score: 0.99 – Model mampu mendeteksi phishing dengan performa 99%.

Skor Keseluruhan:

1. Accuracy: 0.99 – Menunjukkan 99% prediksi pada data uji diklasifikasikan dengan benar oleh model.
2. Macro Average: 0.99 – Rata-rata unweighted dari performa tiap kelas menunjukkan hasil yang merata.
3. Weighted Average: 0,99 – Rata-rata berbobot berdasarkan jumlah sampel pada tiap kelas juga memperlihatkan hasil maksimal.

Dengan demikian, meskipun terdapat sedikit false positive, model masih dapat dikategorikan sangat akurat dan andal dalam mendeteksi email phishing maupun non-phishing berdasarkan data uji yang digunakan.

Setelah model dievaluasi menggunakan metrik klasifikasi, dilakukan analisis lebih lanjut menggunakan confusion matrix. Matriks ini memberikan gambaran visual mengenai jumlah prediksi yang benar dan salah berdasarkan kategori kelas, yaitu Phishing dan Non-Phishing. Berdasarkan hasil pengujian model terhadap data uji sebanyak 11.747 email, diperoleh confusion matrix sebagai berikut:



Gambar. 7 Laporan Confusion Matrix

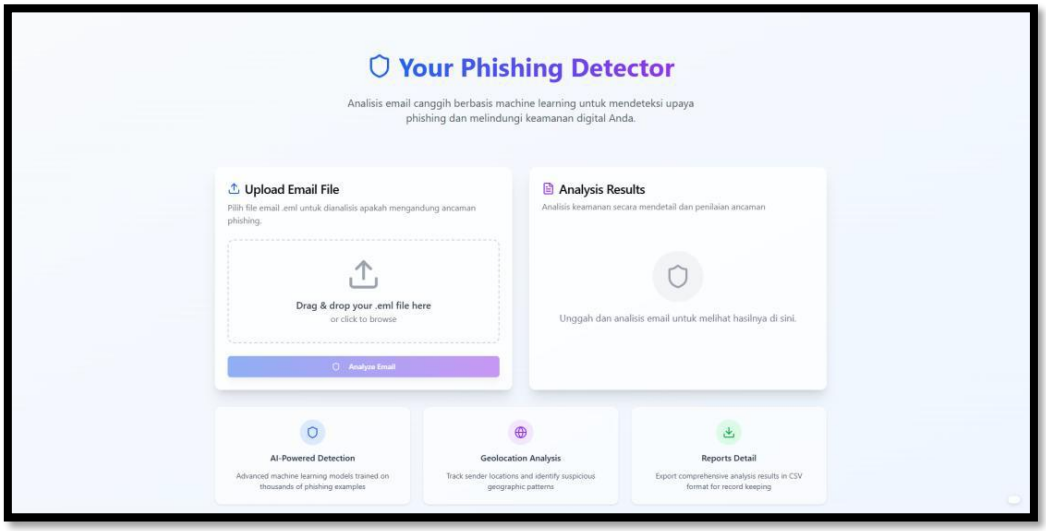
Berikut interpretasi nilai-nilai pada confusion matrix:

1. True Negative (5194): Sebanyak 5.194 email non-phishing (legit) berhasil diklasifikasikan dengan benar sebagai non-phishing.
2. False Positive (0): Tidak ada email non-phishing yang salah diklasifikasikan sebagai phishing.
3. False Negative (27): Sebanyak 27 email phishing gagal dikenali dan justru diklasifikasikan sebagai non-phishing.
4. True Positive (6526): Sebanyak 6.526 email phishing berhasil diklasifikasikan dengan benar sebagai phishing.

Meskipun jumlah kesalahan klasifikasi sangat kecil (hanya 27 dari 11.747 data), keberadaan nilai false negative tetap penting untuk diperhatikan, karena berisiko melewatkan ancaman sebenarnya. Hal ini menunjukkan bahwa model memiliki performa sangat baik, namun tetap memiliki ruang penyempurnaan, terutama jika dihadapkan pada jenis email yang lebih kompleks atau tidak umum.

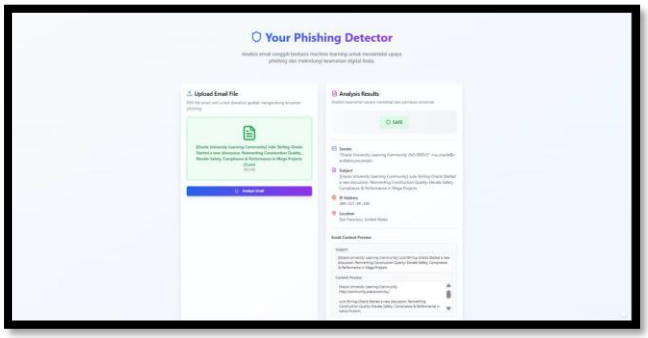
E-ISSN : 2985-4318
P-ISSN : 2985-5772

Selanjutnya model diimplementasikan ke bentuk halaman web sederhana dengan tampilan antarmuka sebagai berikut:



Gambar. 8 Tampilan Awal Aplikasi Email Phishing Detector

Tampilan awal sistem menampilkan tombol unggah file beserta petunjuk singkat. Setelah file dipilih dan diproses, hasil klasifikasi ditampilkan secara otomatis. Untuk memperjelas status email, sistem menampilkan hasil prediksi dalam warna merah untuk email phishing dan warna hijau untuk email yang aman.



Gambar. 9 Output Email Safe/Non Phising

Gambar. 10 Output Email Phising

KESIMPULAN

Nilai akurasi total 99% pada pengujian ini menunjukkan bahwa model Random Forest yang digunakan memiliki performa yang konsisten, meskipun terdapat potensi peningkatan di deteksi phishing yang lebih email resmi. Kinerja pada data non-phishing menunjukkan akurasi sempurna (100%). Hal ini mengindikasikan bahwa model mampu mengenali email yang aman dengan sangat baik. Kinerja pada data phishing hanya (99%) karena terdapat 27 email phishing yang salah diklasifikasikan sebagai non-phishing (false negative). Kesalahan ini dapat disebabkan oleh kemiripan struktur atau kata kunci email phishing dengan email resmi (legitimate), sehingga model kesulitan membedakannya.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

DAFTAR PUSTAKA

[1] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing Attacks : A Recent Comprehensive Study and a New Anatomy," vol. 3, no. March, pp. 1–23, 2021, doi: 10.3389/fcomp.2021.563060.
[2] I. Anti-Phishing Working Group, "Phishing Activity Trends Reports." [Online]. Available: https://apwg.org/trendsreports
[3] T. Langan, *Center, Internet Crime Complain*. 2022.
[4] L. E. O. Breiman, "Random Forests 'Machine Learning,'" pp. 5–32, 2001.
[5] P. P. Prabha, "An Optimized Bagging Learning with Ensemble Feature Selection Method for URL Phishing Detection," vol. 19, pp. 1881–1889, 2024, doi: doi.org/10.1007/s42835-023-01680-z.
[6] A. Pradana, "Jurnal Teknologi Informasi dan Multimedia Implementasi Model Machine Learning untuk Deteksi Phishing dengan Pendekatan Ekstraksi Fitur yang Dioptimalkan," vol. 8, no. 1, pp. 27–40, 2026.
[7] S. K. Ahmad, B. A. Dapshima, and Y. C. Essa, "DETECTION OF PHISHING ATTACKS USING MACHINE LEARNING TECHNIQUES," no. 07, pp. 1166–1176, 2024.
[8] D. Setiawan, "Phishing detection system using machine learning classifiers," 2020.
[9] E. Sangra, R. Agrawal, P. R. Gundalwar, K. Sharma, D. Bangri, and D. Nandi, "Malicious Website Detection Using Random Forest and Pearson Correlation for Effective Feature Selection," vol. 15, no. 8, pp. 772–780, 2024.
[10] P. Resnick, "Internet message format," p. https://tools.ietf.org/HTML/rfc5322, 2008.
[11] U. K. Immanuel, U. K. Immanuel, U. K. Immanuel, and M. T. Elsa, "Analisis perbincangan dalam grup whatsapp dengan k-means clustering," vol. 9, no. 4, pp. 1960–1971, 2024.
[12] A. A. Akinyelu and A. O. Adewumi, "Classification of Phishing Email Using Random Forest Machine Learning Technique," vol. 2014, 2014, doi: 10.1155/2014/425731.
[13] S. Developers, "Scikit-learn Documentation," p. https://scikit-learn.org/dev/versions.html, 2023, [Online]. Available: https://scikit-learn.org/dev/versions.html
[14] IMB, "Supervised Learning." [Online]. Available: https://www.ibm.com/id-id/think/topics/supervised-learning
[15] E. E. H. Lastdrager, "Achieving a consensual definition of phishing based on a systematic review of the literature," vol. 1996, pp. 1–10, 2014.