Jurnal Minfo Polgan (JMP)
Volume 15, Nomor 1, April 2026
DOI : https://doi.org/10.33395/jmp.v15i1.16104
e-ISSN : 2797-3298
p-ISSN : 2089-9424

Perbandingan Algoritma *Machine Learning* dalam Klasifikasi Serangan *Phishing* pada E-mail

[1*]Kiki Putriani Siregar, [2]Budi Triandi, [3]Roslina
[1 2 3] Ilmu Komputer, Fakultas Teknik dan ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia
**Korespondensi*: Putrianikiki1@gmail.com

Submit : 13 Mei 2026 | **Diterima** : 16 April 2026 | **Terbit** : 22 April 2026

ABSTRACT

Phishing attacks via email are a growing cyber threat and have the potential to lead to sensitive data leaks. This study aims to compare the performance of three machine learning algorithms: Gradient Boosting, Random Forest, and Logistic Regression, in classifying phishing and non-phishing emails. The dataset was processed through text cleaning and feature extraction using TF-IDF, then evaluated based on accuracy, precision, recall, and F1-score metrics. The results show that ensemble learning-based algorithms, specifically Gradient Boosting and Random Forest, provide higher performance than Logistic Regression in detecting complex phishing patterns. Although Logistic Regression is simpler and easier to interpret, ensemble models proved more effective in improving classification accuracy.

Keywords: : *Phishing Email, Machine Learning, Gradient Boosting, Random Forest, Logistic Regression.*

ABSTRAK

Serangan phishing melalui email merupakan ancaman siber yang terus meningkat dan berpotensi menyebabkan kebocoran data sensitif. Penelitian ini bertujuan membandingkan kinerja tiga algoritma *Machine Learning*, yaitu Gradient Boosting, Random Forest, dan Logistic Regression dalam mengklasifikasikan email phishing dan non-phishing. Dataset diproses melalui tahapan pembersihan teks dan ekstraksi fitur menggunakan TF-IDF, kemudian dievaluasi berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma berbasis *ensemble learning*, khususnya Gradient Boosting dan Random Forest, memberikan performa lebih tinggi dibandingkan Logistic Regression dalam mendeteksi pola phishing yang kompleks. Meskipun Logistic Regression lebih sederhana dan mudah diinterpretasikan, model ensemble terbukti lebih efektif dalam meningkatkan akurasi klasifikasi

Kata Kunci: Phishing Email, Machine Learning, Gradient Boosting, Random Forest, Logistic Regression.

PENDAHULUAN

Di era digital modern, internet telah menjadi bagian yang tidak terpisahkan dari aktivitas masyarakat, di mana berbagai bentuk komunikasi dan transaksi kini dilakukan secara daring. Pemanfaatan internet telah menjangkau hampir seluruh sektor, mulai dari perdagangan elektronik (*e-commerce*), transportasi, pariwisata, layanan kesehatan, pemerintahan berbasis digital (*e-government*), hingga industri keuangan. Kehadiran internet memungkinkan masyarakat Indonesia untuk mengakses beragam informasi dan layanan melalui berbagai situs web, yang kini berperan sebagai sarana utama dalam memperoleh informasi serta berinteraksi dalam kehidupan sehari-hari.[1]

Serangan phishing telah menyebabkan kerugian finansial yang tidak sedikit serta kebocoran data pribadi, baik bagi individu maupun organisasi. Website sendiri merupakan kumpulan halaman web yang berada dalam satu domain dan biasanya dibuat untuk tujuan atau topik tertentu. Melalui website, pengguna dapat melakukan berbagai aktivitas seperti berbelanja online, membaca berita, mengikuti pembelajaran daring, hingga menjalankan transaksi bisnis. terdapat risiko kejahatan siber, salah satunya adalah phishing website. Phishing website merupakan situs palsu yang dirancang menyerupai tampilan website resmi dengan tujuan

menipu pengguna agar memasukkan informasi pribadinya dan masih banyak pengguna yang tidak menyadari perbedaan antara situs asli dan situs tiruan, sehingga akhirnya menjadi korban. Informasi yang paling sering menjadi sasaran dalam serangan ini adalah username, password, dan nomor PIN, yang kemudian dapat disalahgunakan untuk tindakan kriminal lebih lanjut [2]

Menurut penelitian terdahulu, Sejumlah penelitian sebelumnya telah membahas klasifikasi phishing dengan memanfaatkan algoritma *machine learning*. Salah satunya adalah penelitian yang dilakukan oleh Putri N dan Wijayanto A yang membandingkan beberapa algoritma *data mining* untuk mendeteksi website phishing. Hasil penelitian tersebut menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan tingkat akurasi mencapai 90,77%, lebih tinggi dibandingkan Naïve Bayes yang memperoleh 82,31%, serta algoritma Decision Tree dan SVM. Tidak hanya unggul dari sisi akurasi, Random Forest juga mencatat nilai presisi tertinggi sebesar 87,90% serta sensitivitas mencapai 95,61%, yang menunjukkan kemampuannya dalam mendeteksi website phishing secara lebih konsisten dan akurat.[3]

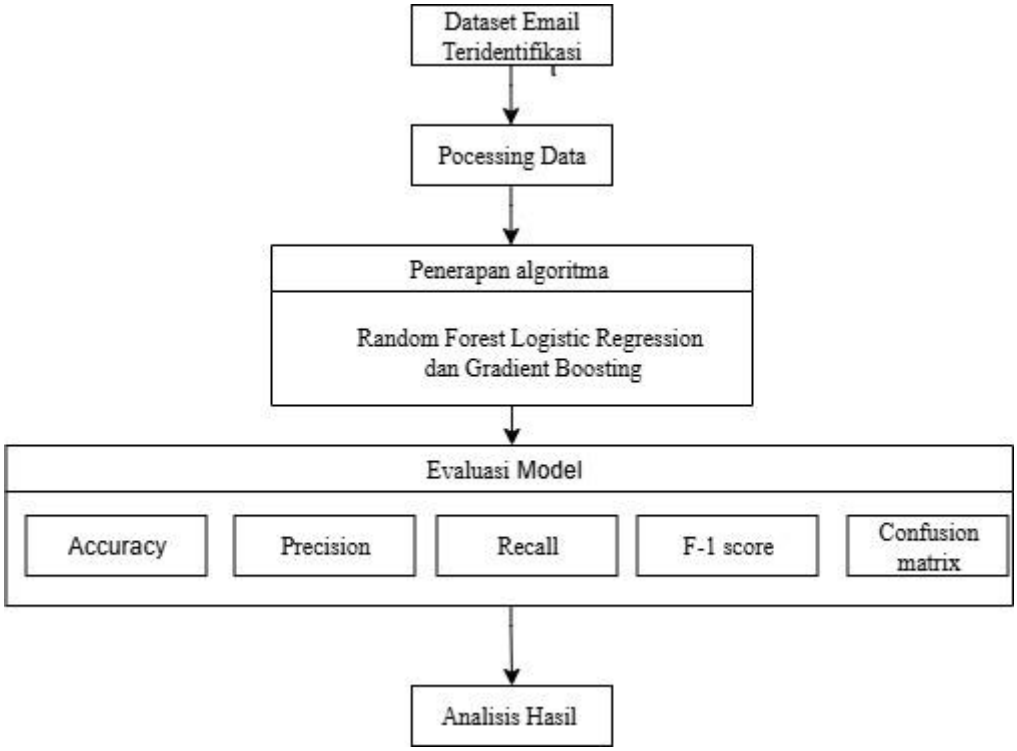
kemudian penelitian yang dilakukan oleh Irawan A, Heryana N, Hopipah H, dan Rahma D membandingkan empat algoritma klasifikasi, yaitu Support Vector Machine (SVM), Decision Tree, Random Forest, dan Multilayer Perceptron (MLP). Berdasarkan hasil pengujian, Multilayer Perceptron menunjukkan kinerja paling unggul dengan tingkat akurasi sebesar 93,15% dan nilai AUC mencapai 0,976. Hasil tersebut menunjukkan bahwa pendekatan berbasis jaringan saraf mampu menangkap pola data phishing dengan tingkat ketepatan yang sangat baik.[4] Sementara itu, penelitian yang dilakukan oleh Mahmud A dan Wirawan S berfokus pada deteksi phishing berdasarkan fitur URL dengan membandingkan tiga algoritma klasifikasi. Hasilnya menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi sebesar 0,834, presisi 0,86, recall 0,83, dan F1-score 0,83. Nilai tersebut sedikit lebih tinggi dibandingkan Decision Tree yang memperoleh akurasi 0,833, serta jauh mengungguli K-Nearest Neighbors yang hanya mencapai akurasi 0,482. Temuan ini semakin memperkuat bahwa Random Forest memiliki kemampuan yang stabil dan konsisten dalam mendeteksi pola phishing.[5]

Berdasarkan penelitian terdahulu yang berfokus pada pengembangan sistem deteksi phishing berbasis web dengan memanfaatkan algoritma Decision Tree. Hasil pengujian menunjukkan bahwa sistem yang dikembangkan mampu mencapai tingkat akurasi sebesar 95,07%, yang menandakan bahwa metode tersebut cukup efektif dalam mengidentifikasi serangan phishing. Hal ini berfokus pada pengembangan sistem deteksi phishing berbasis web dengan memanfaatkan algoritma Decision Tree. Hasil pengujian menunjukkan bahwa sistem yang dikembangkan mampu mencapai tingkat akurasi sebesar 95,07%, yang menandakan bahwa metode tersebut cukup efektif dalam mengidentifikasi serangan phishing.[6]

Penelitian ini berfokus pada metode gradiend boosting , random forest dan Logistic Regression untuk mengklasifikasikan phishing website, berbeda dengan penelitian sebeleumnya yang lebih banyak menggunakan metode Naïve Bayes, KNN DAN SVM. Dari segi objek penelitian ini membandingkan tiga kategori metode beda dengan penelitian sebelumnya yang hanya berfokus pada pengembangan sistem deteksi phishing berbasis web dengan memanfaatkan algoritma Decision Tree. Penelitian ini mempertimbangkan akurasi, precision, recall, F1-score, hingga memakai kurva ROC. Tujuan utama dari penelitian ini adalah untuk membanding metode algoritma mechine learning pada serangan phising e-mail, untuk mengembangkan sistem deteksi phishing yang lebih akurat dan efisien. Selain itu, penelitian ini juga bertujuan untuk membandingkan hasil yang diperoleh dengan metode deteksi lainnya guna menentukan pendekatan terbaik dalam mengidentifikasi situs phishing.

METODE PENELITIAN

Pada penelitian ini dilakukan metode analisis dan teknik perbandingan algoritma klasifikasi menggunakan algoritma Random Forest, Logistic Regression dan Gradient Boosting pada data set email phishing. Penelitian ini bertujuan untuk mengetahui kinerja dari masing-masing algoritma yang di uji dan dapat dilihat dengan tinggak akurasi, presisi, *recall*, dan *f1-score* dalam mengklasifikasikan *email phishing* dan *non phishing.*



Gambar 1. Tahap diagram metode penelitian

tahap ini dilakukan dengan mengumpulkan dan menganalisis penelitian terdahulu. Data yang digunakan dalam penelitian ini dapat diperoleh dari dataset publik seperti Kaggle. Kemudian dilakukan pre-processing data yang bertujuan agar dataset yang digunakan dalam penelitian memiliki kualitas yang baik, konsisten, dan siap dipakai dalam proses pemodelan. Tahapan preprocessing pada model klasifikasi serangan phishing ini memiliki 5 tahapan yaitu normalisasi data, seleksi fitur, Smote, pelabelan dan pembagian data set, Penerapan algoritma data menggunakan 3 algoritma *machine learning* yang dipercaya dapat memberikan hasil yang optimal dalam proses model klasifikasi dataset serangan phishing pada email, sebagaimana dalam pengujian peneliti sebelumnya yang menggunakan *machine learning* dalam teknik klasifikasi. Selanjutnya tahap evaluasi Setelah model machine learning selesai dilatih untuk mengklasifikasikan email phishing, tahapan evaluasi ini penting untuk mengetahui seberapa baik model mampu membedakan antara email phishing dan email yang sah. Beberapa metrik umum yang digunakan adalah akurasi, presisi, recall, F1-score, dan confusion matrix. Kemudian yang terakhir Analisis hasil merupakan proses penafsiran terhadap data yang diperoleh dari pengujian model untuk mengetahui sejauh mana metode yang digunakan mampu menyelesaikan permasalahan penelitian. Dalam penelitian ini, analisis hasil dilakukan dengan mengevaluasi kinerja algoritma Random Forest dengan Logistic Regression dan Gradient Boosting dalam mengklasifikasikan email phishing dan non-phishing, kemudian membandingkan performa keduanya berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

HASIL DAN PEMBAHASAN

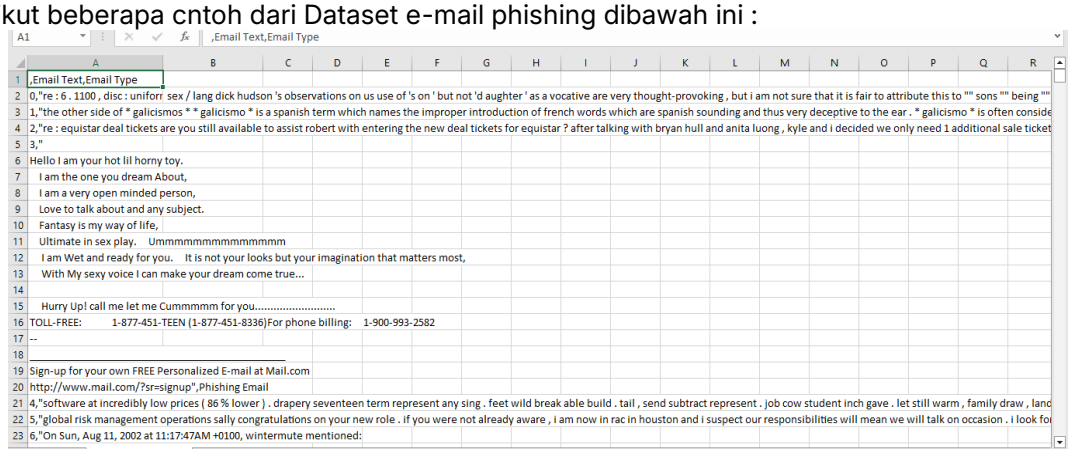
Pada bagian ini, dijelaskan hasil yang diperoleh dari pengujian model klasifikasi terhadap dataset yang telah diproses. Hasil yang akan dibahas meliputi distribusi dataset, kinerja algoritma klasifikasi, dan perbandingan algoritma.

Distribusi Dataset

Dataset yang digunakan dari penelitian ini sebanyak (18.650, 3), yang menunjukkan bahwa dataset terdiri dari 18.650 baris data dan 3 kolom. Hasil ini menandakan bahwa data telah berhasil dimuat secara utuh dan siap untuk diproses. Proses pengambilan data dilakukan dengan cara mengunduh dataset dari sumber-sumber tersebut secara berkala dan menyimpannya dalam format CSV untuk kemudian diproses lebih lanjut. Setelah data

dikumpulkan, dilakukan pre-processing data agar dataset dapat digunakan untuk pelatihan model.

Berikut beberapa cntoh dari Dataset e-mail phishing dibawah ini :



Gambar 2 Dataset e-mail phishing

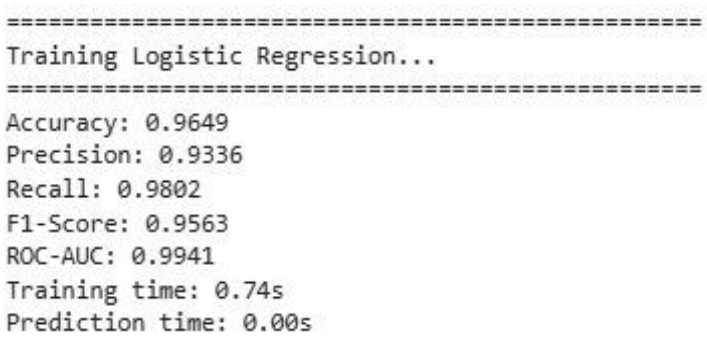
Kinerja algoritma klasifikasi

Dalam analisis performa model klasifikasi, dilakukan pengujian menggunakan beberapa algoritma yang berbeda untuk menentukan tingkat akurasi serta efektivitas model dalam melakukan prediksi terhadap dataset yang digunakan. Pengujian ini bertujuan untuk mengidentifikasi algoritma mana yang memberikan hasil terbaik berdasarkan metrik evaluasi tertentu. Adapun metrik yang digunakan meliputi precision, recall, F1 score. Masing-masing metrik memiliki fungsi tertentu dan memberikan wawasan yang berbeda tentang hasil klasifikasi.

Tahap pre-processing data dalam penelitian ini dilakukan untuk memastikan dataset siap digunakan dalam pelatihan model klasifikasi. Proses ini dimulai dengan dilakukan proses pemuatan dataset email phishing menggunakan fungsi pd.read_csv(). Setelah dataset dimuat, dilakukan pengecekan awal dengan menampilkan ukuran dataset menggunakan atribut df.shape, penghapusan data yang memiliki nilai kosong (*missing value*) menggunakan perintah df.dropna(), proses pembentukan variabel kelas baru dengan nama label menggunakan fungsi apply() pada kolom *Email Type,* dilakukan perhitungan distribusi data berdasarkan jenis email yang terdapat pada kolom *Email Type* menggunakan fungsi value_counts(). Selanjutnya pada tahapan visualisasi data, dilakukan pembuatan grafik batang (*bar chart*) untuk menggambarkan distribusi jenis email berdasarkan kolom *Email Type*. Visualisasi ini dibuat menggunakan library Matplotlib dengan pengaturan ukuran gambar sebesar *(12, 6)* agar grafik dapat ditampilkan dengan jelas dan mudah dibaca. Selanjutnya dilakukan penambahan label nilai pada setiap batang grafik menggunakan perulangan for yang mengakses objek *bar*, Kemudian dilakukan pembuatan diagram lingkaran (*pie chart*) untuk menampilkan persentase distribusi jenis email berdasarkan kolom *Email Type*. Selanjutnya dilakukan pembersihan data yaiu membersihkan dan memproses tek e-mail, mengubah menjadi huruf kecil. Selaanjutnya dilakukan praposes teks. Tahap berikutnya adalah normalisasi data, di mana semua teks URL diubah menjadi huruf kecil dan data kategorikal dikonversi ke format numerik agar dapat digunakan dalam model machine learning. Terakhir, dataset dibagi menjadi training set (80%) untuk melatih model dan testing set (20%) untuk menguji performa model.

Hasil akurasi pengujian menggunakan algoritma Logistic Regression,

Algoritma Logistic Regression pada penelitian ini digunakan untuk mengembangkan model klasifikasi yang mampu memperkirakan apakah suatu email termasuk dalam kategori phishing atau bukan. Melalui analisis terhadap berbagai fitur yang terdapat pada email, algoritma ini dapat menghitung kemungkinan suatu pesan memiliki karakteristik serangan phishing. Hasil kurasi algoritma Logistic Regression dapat ditunjukkan pada gambar 3.



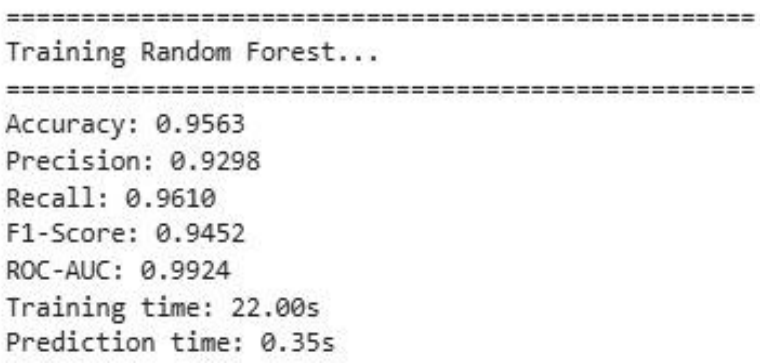
Gambar 3. Hasil Akurasi Pengujian Algoritma Logistic Regression

algoritma Logistic Regression, diperoleh nilai akurasi sebesar 0.9649%, yang menunjukkan bahwa model mampu mengklasifikasikan email phishing dan non-phishing dengan tingkat ketepatan yang sangat tinggi. Nilai presisi sebesar 0.9336% mengindikasikan bahwa sebagian besar email yang diprediksi sebagai phishing benar-benar merupakan email phishing, sehingga tingkat kesalahan *false positive* relatif rendah. Sementara itu, nilai recall sebesar 0.9802% menunjukkan kemampuan model yang sangat baik dalam mendeteksi hampir seluruh email phishing yang ada pada data uji.sedangkan presisi dan recall tersebut menghasilkan F1-score sebesar 0.9563%, Nilai ROC-AUC sebesar 0.9941% yang mencerminkan keseimbangan kinerja model dalam mendeteksi email phishing secara akurat dan konsisten. Hasil pengujian menunjukkan bahwa waktu pelatihan Logistic Regression hanya sebesar 0,74 detik, sedangkan waktu prediksi mendekati 0.00 detik, yang menandakan bahwa model ini sangat efisien dan cepat dalam melakukan inferensi.

Sehingga secara keseluruhan bahwa algoritma Logistic Regression efektif dan efisien digunakan sebagai model klasifikasi dalam deteksi serangan phishing pada email, karena mampu menghasilkan tingkat akurasi yang tinggi sekaligus memiliki waktu komputasi yang cepat dalam proses pelatihan maupun prediksi.

Hasil akurasi pengujian algoritma Random Forest

Setelah dilakukan pelatihan dan pengujian menggunakan algoritma Random Forest , diperoleh hasil evaluasi model berdasarkan classification report ditunjukkan pada gambar 4.



Gambar 4. Hasil akurasi pengujian algoritma Random Forest

Pengujian menggunakan algoritma Random Forest, diperoleh nilai akurasi sebesar 0.9563%, yang menunjukkan bahwa model mampu mengklasifikasikan email phishing dan non-phishing dengan tingkat ketepatan yang tinggi. Nilai presisi sebesar0. 9298% dan recall sebesar 0.9610% mengindikasikan bahwa model memiliki kemampuan yang seimbang dalam mengidentifikasi email phishing secara tepat sekaligus mendeteksi sebagian besar email phishing yang sebenarnya. Keseimbangan antara presisi dan recall tersebut tercermin dari nilai F1-score sebesar 0.9452%, yang menunjukkan performa model yang stabil dan konsisten dalam proses klasifikasi. Nilai ROC-AUC sebesar 0.9924% menunjukkan bahwa Random Forest memiliki kemampuan yang sangat baik dalam membedakan email phishing dan non-phishing

pada berbagai ambang keputusan. Namun, jika dibandingkan dengan Logistic Regression, Random Forest memerlukan waktu pelatihan yang jauh lebih lama, yaitu 22.00 detik, serta waktu prediksi sebesar 0,35 detik. Hal ini disebabkan oleh kompleksitas model yang membangun banyak pohon keputusan secara paralel.

algoritma Random Forest terbukti mampu memberikan kinerja yang sangat baik dalam membangun model klasifikasi dengan tingkat akurasi dan kemampuan deteksi phishing yang tinggi. Meskipun demikian, proses pelatihan dan prediksi pada algoritma ini membutuhkan waktu komputasi yang lebih besar dibandingkan metode yang lebih sederhana seperti Logistic Regression. Hal ini terjadi karena Random Forest membangun banyak pohon keputusan untuk menghasilkan prediksi yang lebih stabil.

Hasil akurasi pengujian algoritma Gradient Boosting,
Selanjutnya dilakukan dengan pengujian menggunakan algoritma Gradient Boosting dan berikut hasil akurasi nya di tunjukkan dibawah ini dengan gambar 5.

```
==================================================
Training Gradient Boosting...
==================================================
Accuracy: 0.9265
Precision: 0.9136
Recall: 0.8974
F1-Score: 0.9055
ROC-AUC: 0.9785
Training time: 72.00s
Prediction time: 0.03s
```

Gambar 5 hasil pengujian akurasi algoritma Gradient Boosting

Pelatihan dan pengujian menggunakan algoritma Gradient Boosting, diperoleh nilai akurasi sebesar 0.9265%, yang menunjukkan bahwa model mampu mengklasifikasikan email phishing dan non-phishing dengan tingkat ketepatan yang cukup tinggi. Nilai presisi sebesar 0.9136% mengindikasikan bahwa sebagian besar email yang diprediksi sebagai phishing memang benar merupakan email phishing, sehingga tingkat kesalahan *false positive* relatif rendah. Namun demikian, nilai recall sebesar 0.8974% menunjukkan bahwa masih terdapat sejumlah email phishing yang tidak berhasil terdeteksi oleh model, yang berdampak pada penurunan kemampuan model dalam menangkap seluruh kasus phishing. Pada nilai F1-score sebesar 0.9055%, yang menunjukkan bahwa keseimbangan antara presisi dan recall pada Gradient Boosting masih berada di bawah dua model lainnya. Meskipun demikian, nilai ROC-AUC sebesar 0.9785% menandakan bahwa model tetap memiliki kemampuan diskriminatif yang sangat baik dalam membedakan email phishing dan non-phishing. Dari sisi efisiensi komputasi, Gradient Boosting membutuhkan waktu pelatihan yang paling lama, yaitu 72,00 detik, meskipun waktu prediksinya relatif cepat 0.03%.

Gradient Boosting mampu menghasilkan model klasifikasi yang cukup baik, tetapi dengan kebutuhan waktu pelatihan yang lebih besar serta performa deteksi yang sedikit lebih rendah dibandingkan model lain.

Pembahasan

Pada bagian ini, hasil penelitian dianalisis dan dikaitkan dengan literatur yang relevan. Pada penelitian ini, tiga algoritma yang sering digunakan dalam klasifikasi data telah diuji dan dibandingkan, yaitu Logistic Regression, Random Forest dan Gradient Boosting. Masing-masing algoritma memiliki pendekatan yang berbeda dalam menangani data serta dalam proses pengambilan keputusan terhadap prediksi kelas suatu objek. Berdasarkan hasil pengujian akurasi dan metrik evaluasi lainnya, akan dilakukan perbandingan performa tiga algoritma klasifikasi tersebut.

Perbandingan Algoritma

Akurasi adalah metrik utama yang digunakan untuk mengukur performa model klasifikasi. Hasil akurasi dari ketiga algoritma ditunjukkan pada tabel 1.

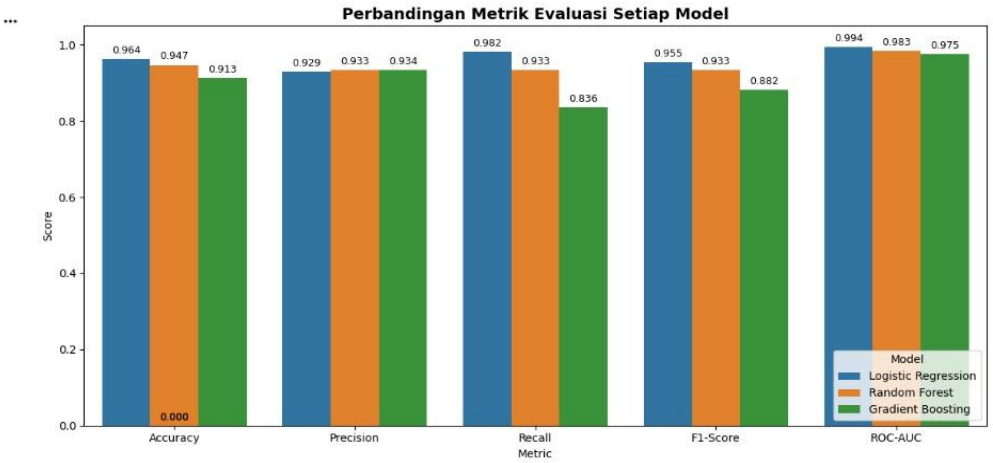
Tabel 4. Perbandingan algoritma

Algoritma	Akurasi
Logistic Regression	0.9649
Random Forest	0.9563
Gradient Boosting	0.9265

Hasil pengujian menunjukkan bahwa Logistic Regression memiliki akurasi sebesar 0.9649 daN algoritma Random Forest memiliki akurasi sebesar 0.9563 sedangkan Gradient Boosting memiliki akurasi sebesar 0.9265. Meskipun logistic Regression unggul dalam kecepatan pemrosesan, algoritma ini mengalami kesulitan dalam mengklasifikasikan kelas tertentu, terutama Phishing.

Perbandingan Matrix Evaluasi

Data hasil evaluasi pada setiap model klasifikasi adalah data awal yang tersimpan dalam format panjang. Sehingga setiap nilai metrik seperti akurasi, precision, recall, dan F1-score dapat dianalisis dan divisualisasikan secara seragam untuk seluruh model. Penggunaan grafik batang dengan sumbu horizontal sebagai jenis metrik dan sumbu vertikal sebagai nilai skor memungkinkan identifikasi perbedaan performa model secara cepat dan intuitif. Pemberian warna yang berbeda untuk setiap model berfungsi untuk menegaskan perbandingan kinerja antar algoritma pada metrik yang sama. Berikut dibawah ini gambar grafik perbandingan Matrix evaluasi



Gambar 6. grafik perbandingan matrix evaluasi

KESIMPULAN

Berdasarkan seluruh proses penelitian yang telah dilakukan, mulai dari pengolahan data hingga evaluasi model, dapat disimpulkan bahwa penerapan algoritma machine learning mampu memberikan hasil yang sangat baik dalam mendeteksi email phishing. Hasil pengujian menunjukkan bahwa ketiga algoritma yang digunakan, yaitu Logistic Regression, Random Forest, dan Gradient Boosting, sama-sama mampu melakukan klasifikasi email phishing dengan tingkat akurasi yang tinggi. Sehingga antara ketiga model tersebut, Logistic Regression menunjukkan performa paling optimal dan konsisten, dengan nilai accuracy sebesar 96,4%, precision sebesar 92,9%, recall sebesar 98,2%, F1-score sebesar 95,5%, serta ROC-AUC sebesar 0,994.

UCAPAN TERIMA KASIH

Terima kasih kepada bapak/ ibu Dosen pembimbing serta bapak Kaprodi yang sudah memberikan saya saran dan motivasi dan terima kasih juga kepada keluarga dan teman saya yang paling saya sayangi.

REFERENSI

Ahmad Turmudi Zy, & M. Makmun Effendi. (2023). KLASIFIKASI EMAIL PHISHING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal RESTIKOM : Riset*

Teknik Informatika dan Komputer, *5*(2), 148 157.https://doi.org/10.52005/restikom.v5i2.152

Ahmadian, H., & Sabri, A. (2021). TEKNIK PENYERANGAN PHISHING PADA SOCIAL ENGINEERING MENGGUNAKAN SET DAN PENCEGAHANNYA. *Djtechno Jurnal Teknologi Informasi*, *2*(1), 13–20. https://doi.org/10.46576/djtechno.v2i1.1251

Alhuzali, A., Alloqmani, A., Aljabri, M., & Alharbi, F. (2025). In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets. *Applied Sciences*, *15*(6), 3396. https://doi.org/10.3390/app15063396

Almomani, O. (2020). A Feature Selection Model for Network Intrusion Detection System Based on PSO, GWO, FFA and GA Algorithms. *Symmetry*, *12*(6), 1046. https://doi.org/10.3390/sym12061046

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. *Clinical Pharmacology & Therapeutics*, *107*(4), 871–885. https://doi.org/10.1002/cpt.1796

Fadli, M., & Saputra, R. A. (2023). *KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE*. *12*(02).

Faruq, A., Khaeruddin, K., & Lestandy, M. (2020). Sistem Keamanan Multi Mail Server dengan Teknik Enkripsi OPENPGP pada Zimbra Exchange Open Source Software. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *7*(3), 493. https://doi.org/10.25126/jtiik.2020731869

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). *A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications* (No. arXiv:2001.06937). arXiv. https://doi.org/10.48550/arXiv.2001.06937

Gumay, B., Hendrawan, A. H., & Kusumah, F. S. F. (2024). ANALISIS DAMPAK ANCAMAN CYBERCRIME TERHADAP DATA MAHASISWA PADA SERANGAN WEB PHISING SIAK UIKA. *INFOTECH journal*, *10*(2), 297–305. https://doi.org/10.31949/infotech.v10i2.11463

Mahmud Nawawi, H., Baitul Hikmah, A., Mustopa, A., & Wijaya, G. (2024). Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir. *Jurnal SAINTEKOM*, *14*(1), 13–25. https://doi.org/10.33020/saintekom.v14i1.512

Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, *11*(1), 59–66. https://doi.org/10.34010/komputika.v11i1.4350

Raihan, A., Fadhli, M., & Lindawati, L. (2024). IMPLEMENTATION OF DEEP LEARNING FOR DETECTING PHISHING ATTACKS ON WEBSITES WITH COMBINATION OF CNN AND LSTM. *Jurnal Teknik Informatika (Jutif)*, *5*(5), 1451–1459. https://doi.org/10.52436/1.jutif.2024.5.5.2446

Tang, L., & Mahmoud, Q. H. (2021). A Survey of Machine Learning-Based Solutions for Phishing Website Detection. *Machine Learning and Knowledge Extraction*, *3*(3), 672–694. https://doi.org/10.3390/make3030034

Zhang, Y., Li, Q., & Xin, Y. (2024). Research on eight machine learning algorithms applicability on different characteristics data sets in medical classification tasks. *Frontiers in Computational Neuroscience*, *18*. https://doi.org/10.3389/fncom.2024.1345575

Ahmad Turmudi Zy, & M. Makmun Effendi. (2023). KLASIFIKASI EMAIL PHISHING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR. *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, *5*(2), 148–157. https://doi.org/10.52005/restikom.v5i2.152

Ahmadian, H., & Sabri, A. (2021). TEKNIK PENYERANGAN PHISHING PADA SOCIAL ENGINEERING MENGGUNAKAN SET DAN PENCEGAHANNYA. *Djtechno Jurnal Teknologi Informasi*, *2*(1), 13–20. https://doi.org/10.46576/djtechno.v2i1.1251

Alhuzali, A., Alloqmani, A., Aljabri, M., & Alharbi, F. (2025). In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets. *Applied Sciences*, *15*(6), 3396. https://doi.org/10.3390/app15063396

Almomani, O. (2020). A Feature Selection Model for Network Intrusion Detection System Based on PSO, GWO, FFA and GA Algorithms. *Symmetry*, *12*(6), 1046. https://doi.org/10.3390/sym12061046

Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning. *Clinical Pharmacology & Therapeutics*, *107*(4), 871–885. https://doi.org/10.1002/cpt.1796

Fadli, M., & Saputra, R. A. (2023). *KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE*. *12*(02).

Faruq, A., Khaeruddin, K., & Lestandy, M. (2020). Sistem Keamanan Multi Mail Server dengan Teknik Enkripsi OPENPGP pada Zimbra Exchange Open Source Software. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *7*(3), 493. https://doi.org/10.25126/jtiik.2020731869

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). *A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications* (No. arXiv:2001.06937). arXiv. https://doi.org/10.48550/arXiv.2001.06937

Gumay, B., Hendrawan, A. H., & Kusumah, F. S. F. (2024). ANALISIS DAMPAK ANCAMAN CYBERCRIME TERHADAP DATA MAHASISWA PADA SERANGAN WEB PHISING SIAK UIKA. *INFOTECH journal*, *10*(2), 297–305. https://doi.org/10.31949/infotech.v10i2.11463

Mahmud Nawawi, H., Baitul Hikmah, A., Mustopa, A., & Wijaya, G. (2024). Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir. *Jurnal SAINTEKOM*, *14*(1), 13–25. https://doi.org/10.33020/saintekom.v14i1.512

Putri, N. B., & Wijayanto, A. W. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, *11*(1), 59–66. https://doi.org/10.34010/komputika.v11i1.4350

Raihan, A., Fadhli, M., & Lindawati, L. (2024). IMPLEMENTATION OF DEEP LEARNING FOR DETECTING PHISHING ATTACKS ON WEBSITES WITH COMBINATION OF CNN AND LSTM. *Jurnal Teknik Informatika (Jutif)*, *5*(5), 1451–1459. https://doi.org/10.52436/1.jutif.2024.5.5.2446

Tang, L., & Mahmoud, Q. H. (2021). A Survey of Machine Learning-Based Solutions for Phishing Website Detection. *Machine Learning and Knowledge Extraction*, *3*(3), 672–694. https://doi.org/10.3390/make3030034

Zhang, Y., Li, Q., & Xin, Y. (2024). Research on eight machine learning algorithms applicability on different characteristics data sets in medical classification tasks. *Frontiers in Computational Neuroscience*, *18*. https://doi.org/10.3389/fncom.2024.1345575