Contents lists available at JESA

Journal of Engineering and Science Application

journal homepage: https://jesa.aks.or.id/index.php/jesa/index

Deteksi Email Phishing Menggunakan *Machine Learning* Dengan Algoritma *Random Forest Classifier* dan *Random Search Cross-Validation*

Andi Muhammad Isa[a], Deanna Durbin Hutagalung[b]

[a]*Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia*

Article Info

Keywords:
Phishing,
Deteksi Email,
Machine Learning,
Random Forest

ABSTRACT

Email is widely used in educational institutions, but its operational convenience is accompanied by phishing threats that can lead to credential theft, data leakage, and other security incidents. This study develops a machine learning-based system for classifying email text into Safe Email and Phishing Email categories using Random Forest optimized with Random Search Cross-Validation (RSCV). The research uses historical email data from SDN Limo 1 Kota Depok. The processing stages include text cleaning, case folding, masking of email addresses and URLs, stopword removal, stemming, outlier handling, and TF-IDF feature extraction. The dataset is divided into 80% training and 20% testing data. Random Forest is evaluated in a standard configuration and after hyperparameter optimization using five-fold cross-validation and ten random iterations. The optimized model achieves 96.99% accuracy, 97% macro precision, 97% macro recall, and 97% macro F1-score, improving accuracy by 2.51 percentage points over the standard model. The model is implemented in a Streamlit web application to support practical preliminary email screening in educational environments. The application also provides prediction probabilities, an explainable analysis of influential terms, testing history, role-based access, and data export for authorized users. The results indicate that hyperparameter optimization can improve classification performance while the integrated application can support faster initial screening before further security verification is performed.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Andi Muhammad Isa
Program Studi Teknik Informatika, Fakultas Ilmu Komputer
Universitas Pamulang
Tangerang Selatan, Indonesia
Email: isandimuh378@gmail.com

1. INTRODUCTION

Penggunaan email sebagai sarana komunikasi dan administrasi di institusi pendidikan telah menjadi bagian penting dari kegiatan operasional. Namun, tingginya intensitas penggunaan email juga membuka peluang terjadinya serangan siber, termasuk phishing yang memanfaatkan rekayasa sosial untuk mendorong penerima mengeklik tautan berbahaya atau memberikan informasi sensitif [1]–[4]. Ancaman tersebut menjadi semakin penting diperhatikan karena proses pemeriksaan pesan yang dilakukan secara manual bergantung pada ketelitian pengguna dan dapat mengalami kesalahan manusia.

Pendekatan machine learning dapat digunakan untuk membantu melakukan penyaringan awal berdasarkan karakteristik teks email. Penelitian terdahulu menunjukkan bahwa Random Forest dapat digunakan untuk klasifikasi email phishing, sedangkan TF-IDF membantu mengubah teks menjadi representasi numerik yang menonjolkan kata-kata informatif [5]–[10]. Studi lain juga memperlihatkan bahwa performa model dapat berbeda menurut algoritma, dataset, dan strategi validasi yang digunakan [17]–[20].

Random Forest memiliki keunggulan karena menggabungkan banyak pohon keputusan sehingga mampu menangani pola fitur yang beragam. Meskipun demikian, kinerja model dipengaruhi oleh konfigurasi hyperparameter. Random Search Cross-Validation (RSCV) dapat digunakan untuk mencari kombinasi parameter yang lebih sesuai melalui evaluasi silang tanpa harus menguji seluruh kombinasi seperti Grid Search [13]–[15]. Pendekatan ini relevan untuk meningkatkan kemampuan generalisasi model pada dataset email yang memiliki variasi bahasa dan pola pesan.

Penelitian ini menggunakan data email historis dari SDN Limo 1 Kota Depok dengan dua kelas, yaitu Safe Email dan Phishing Email. Kebaruan penelitian terletak pada pengujian Random Forest standar dan Random Forest yang dioptimasi menggunakan RSCV pada data email nyata dari lingkungan sekolah, kemudian mengimplementasikan model terbaik ke dalam aplikasi berbasis Streamlit. Tujuan penelitian adalah menghasilkan model deteksi yang memiliki performa lebih baik sekaligus menyediakan antarmuka yang dapat digunakan untuk pemeriksaan awal email.

Hasil penelitian diharapkan dapat memberikan dukungan praktis bagi staf administrasi dan pengelola sekolah dalam mengenali pesan yang berpotensi phishing. Sistem tidak dimaksudkan untuk menggantikan pemeriksaan keamanan secara menyeluruh, tetapi menjadi alat bantu awal yang dapat mempercepat proses penyaringan dan memberikan informasi probabilitas hasil klasifikasi.

Penelitian ini memiliki tujuan utama untuk membandingkan performa Random Forest standar dengan Random Forest yang dioptimasi menggunakan Random Search Cross-Validation serta mengimplementasikan model terbaik ke dalam aplikasi deteksi email. Fokus tersebut penting karena model dengan akurasi tinggi tetap perlu diuji dalam bentuk aplikasi yang dapat digunakan oleh pengguna nonteknis. Dengan adanya antarmuka berbasis web, proses klasifikasi dapat dilakukan tanpa menjalankan kode pemrograman secara langsung. Hasil prediksi juga dapat disajikan bersama probabilitas dan informasi pendukung sehingga pengguna memperoleh gambaran mengenai tingkat keyakinan model.

Selain persoalan teknis pada model klasifikasi, konteks penggunaan email di lingkungan pendidikan perlu diperhatikan karena pesan yang masuk tidak seluruhnya memiliki pola yang sama. Email administrasi, pemberitahuan kegiatan, komunikasi guru, informasi orang tua, dan pesan promosi dapat menggunakan kosakata yang beragam. Kondisi tersebut menyebabkan sistem deteksi tidak cukup hanya mengenali satu atau dua kata yang dianggap mencurigakan. Model perlu mempelajari kombinasi karakteristik teks agar keputusan klasifikasi tidak terlalu bergantung pada kata tertentu. Oleh karena itu, penelitian ini memanfaatkan representasi TF-IDF dan algoritma Random Forest untuk menangkap hubungan antara pola kata dan label email secara lebih sistematis.

2. RESEARCH METHOD

Penelitian menggunakan pendekatan kuantitatif komputasional dengan tahapan pengumpulan data, pra-pemrosesan teks, ekstraksi fitur, pembagian dataset, pemodelan, optimasi hyperparameter, evaluasi, dan implementasi aplikasi. Alur penelitian disusun agar setiap tahap dapat dievaluasi secara terukur dan hasil model dapat dibandingkan antara konfigurasi standar dan konfigurasi hasil optimasi.

2.1 Data Acquisition and Preprocessing

Dataset berupa email historis operasional dari SDN Limo 1 Kota Depok. Data awal berjumlah 5.075 email dan setelah proses pembersihan serta penyaringan diperoleh 2.989 email yang digunakan dalam pemodelan. Data diberi label Safe Email sebagai kelas 0 dan Phishing Email sebagai kelas 1. Tahap pra-pemrosesan meliputi penghapusan nilai kosong dan duplikat, case folding, masking alamat email dan URL menggunakan pola Regular Expression, penghapusan tanda baca, stopword removal, serta stemming menggunakan pustaka Sastrawi.

Untuk mengurangi pengaruh data yang terlalu pendek atau terlalu panjang, diterapkan penyaringan berdasarkan panjang teks. Email dengan panjang kurang dari 10 karakter atau lebih dari 5.000 karakter dikeluarkan dari proses pemodelan. Selanjutnya dataset dibagi menjadi 80% data latih dan 20% data uji. Hasil pembagian akhir digunakan secara konsisten untuk membandingkan Random Forest standar dengan model yang dioptimasi.

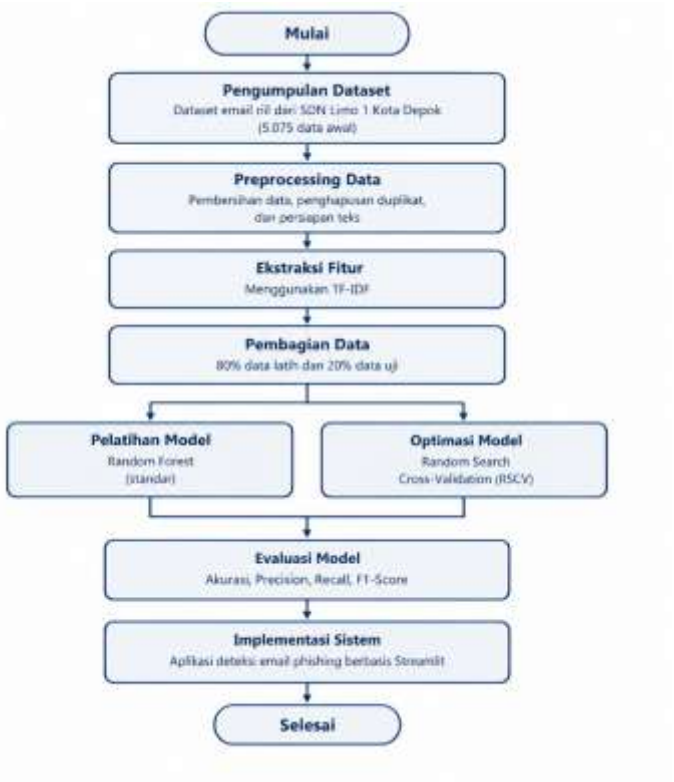


Figure 1. Research workflow for phishing email detection

2.2 TF-IDF Feature Extraction

Teks yang telah dibersihkan ditransformasikan menjadi fitur numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Konfigurasi menggunakan maksimal 1.000 fitur, n-gram unigram dan bigram, min_df = 5, dan max_df = 0,7. Secara umum, bobot TF-IDF dapat dinyatakan sebagai perkalian antara frekuensi suatu term pada dokumen dan faktor inverse document frequency. Representasi ini digunakan agar kata yang terlalu umum memiliki bobot lebih kecil, sedangkan kata yang lebih informatif bagi pembeda kelas memiliki bobot lebih besar.

Analisis karakteristik teks menunjukkan bahwa pesan phishing pada dataset cenderung memuat kosakata yang mendorong tindakan, seperti “link”, “klik disini”, “menang”, dan “hadiah”. Sementara itu, email aman lebih banyak memuat istilah administratif seperti “rapat”, “jadwal”, dan “absensi”. Pola tersebut digunakan sebagai gambaran karakteristik data, bukan sebagai aturan tunggal dalam menentukan kelas.



Figure 2. Word cloud characteristics of Safe Email and Phishing Email

2.3 Random Forest and RSCV

Random Forest digunakan sebagai algoritma klasifikasi utama. Model standar digunakan sebagai baseline, kemudian dibandingkan dengan Random Forest yang dioptimasi menggunakan Random Search Cross-Validation. RSCV menerapkan 5-fold cross-validation dan 10 iterasi pencarian acak. Parameter yang dicari meliputi n_estimators, max_depth, min_samples_split, min_samples_leaf, dan bootstrap. Pembobotan class_weight='balanced' digunakan pada model optimasi untuk membantu menangani perbedaan jumlah sampel antar kelas.

Parameter	RF Standar	Ruang Pencarian RSCV
n_estimators	100	100, 200, 300
max_depth	10	5, 10
min_samples_split	2	5, 10
min_samples_leaf	5	10, 20
bootstrap	True	True
class_weight	None	Balanced
random_state	42	42

Table 1. Random Forest configuration and RSCV search space

2.4 Evaluation

Evaluasi model menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Selain evaluasi model, aplikasi diuji menggunakan Black Box Testing untuk memastikan fungsi autentikasi, klasifikasi, pencatatan riwayat, dan pembatasan hak akses berjalan sesuai kebutuhan. Evaluasi ini membedakan performa model secara statistik dari keberfungsian aplikasi sebagai produk perangkat lunak.

3. RESULTS AND DISCUSSION

3.1 Dataset Characteristics

Setelah tahap preprocessing, dataset yang digunakan terdiri atas 2.989 email. Kelas Safe Email berjumlah 1.637 data atau 54,80%, sedangkan Phishing Email berjumlah 1.352 data atau 45,20%. Komposisi tersebut menunjukkan bahwa distribusi kelas tidak sepenuhnya seimbang, tetapi perbedaannya masih dapat ditangani melalui pembobotan kelas pada proses optimasi.

Email Class	Samples	Percentage
Phishing	1,352	45.20%
Safe Email	1,637	54.80%
Total	2,989	100%

Table 2. Final dataset distribution

3.2 Model Performance

Random Forest standar menghasilkan akurasi 94,48% dan digunakan sebagai baseline. Setelah optimasi menggunakan RSCV, akurasi meningkat menjadi 96,99%. Peningkatan sebesar 2,51 poin persentase menunjukkan bahwa pencarian kombinasi hyperparameter memberikan konfigurasi yang lebih sesuai dengan karakteristik dataset dibandingkan parameter standar.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest Standard	94.48%	95%	95%	94%
Random Forest + RSCV	96.99%	97%	97%	97%

Table 3. Comparison of classification performance

Hasil tersebut sejalan dengan penelitian yang menunjukkan bahwa Random Forest dan pendekatan berbasis TF-IDF dapat memberikan performa kompetitif pada klasifikasi email phishing [9], [17], [22]. Perbedaan hasil antarpenelitian tetap perlu dilihat berdasarkan dataset, pembagian data, preprocessing, dan konfigurasi eksperimen yang digunakan.

3.3 Confusion Matrix Analysis

Pada model Random Forest standar, sebanyak 296 Safe Email diklasifikasikan dengan benar dan 32 Safe Email salah diklasifikasikan sebagai phishing. Untuk kelas phishing, 269 data berhasil dikenali dan satu data salah diklasifikasikan sebagai aman. Setelah optimasi RSCV, jumlah Safe Email yang benar meningkat menjadi 311, sedangkan kesalahan Safe Email sebagai phishing turun menjadi 17. Kelas phishing tetap memiliki 269 prediksi benar dan satu false negative.

Actual / Predicted	Safe Email	Phishing Email
Safe Email	311	17
Phishing Email	1	269

Table 4. Confusion matrix of Random Forest + RSCV

Dengan demikian, jumlah prediksi benar meningkat dari 565 menjadi 580 data dari total 598 data pengujian. Pada tingkat kelas, Safe Email memperoleh precision 100%, recall 95%, dan F1-score 97%, sedangkan

Phishing Email memperoleh precision 94%, recall 100%, dan F1-score 97%. Recall 100% pada kelas phishing menunjukkan bahwa seluruh data phishing pada data uji berhasil dikenali oleh model.

3.4 System Implementation

Model terbaik kemudian diimplementasikan ke dalam aplikasi berbasis Streamlit. Halaman utama menyediakan input teks email dan tombol untuk memulai proses deteksi. Sistem menampilkan kelas hasil prediksi beserta probabilitas Safe Email dan Phishing Email sehingga pengguna memperoleh gambaran tingkat keyakinan model.

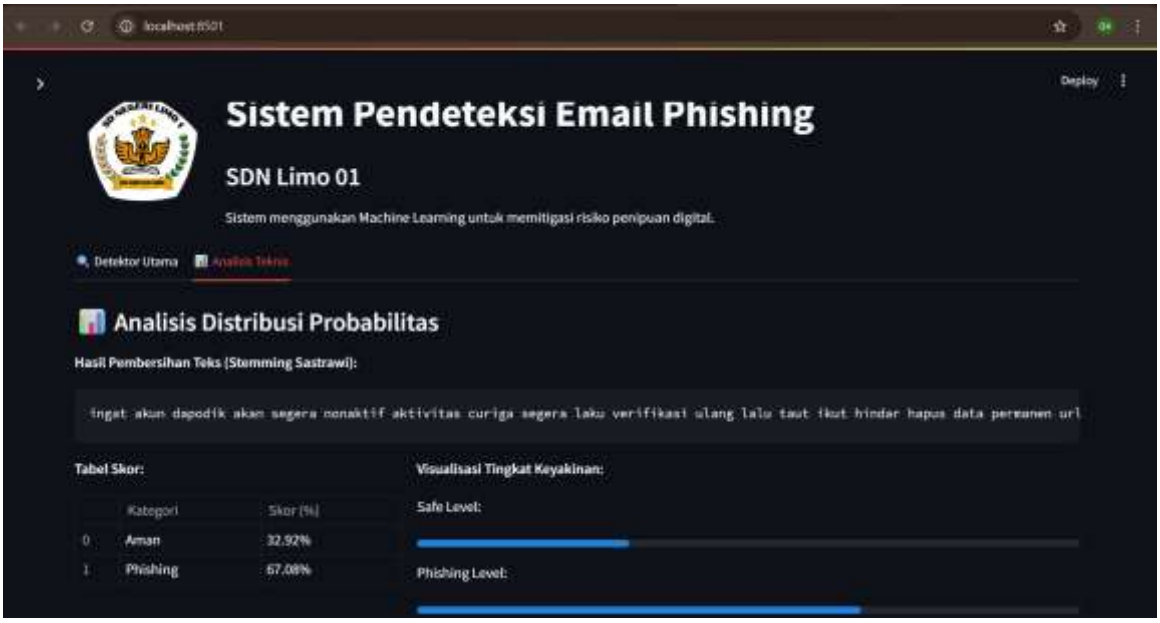


Figure 3. Main interface and probability analysis of the phishing detection system

Aplikasi juga menyediakan analisis teknis berbasis Explainable Artificial Intelligence (XAI) yang menampilkan kata-kata yang memiliki pengaruh terhadap hasil klasifikasi. Fitur ini membantu pengguna memahami bahwa keputusan model berasal dari pola fitur teks, bukan sekadar label akhir.

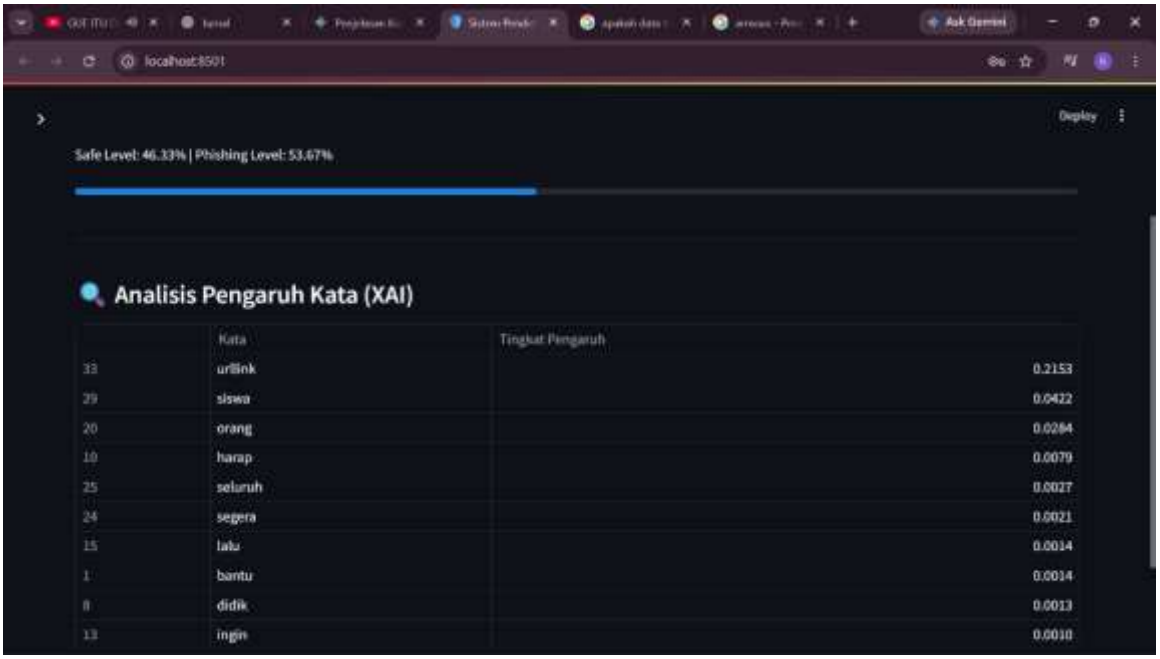


Figure 4. Technical analysis using XAI

Selain fungsi deteksi, sistem menyimpan riwayat pengujian yang memuat waktu, aktor, teks email, prediksi, dan skor probabilitas. Hak akses dibedakan antara Staf TU dan Kepala Sekolah. Staf TU dapat melihat riwayat, sedangkan Kepala Sekolah memperoleh fungsi tambahan untuk mengunduh data riwayat dalam format CSV.



Figure 5. School email detection history and access control

3.5 Functional Testing

Black Box Testing dilakukan terhadap 10 skenario yang mencakup login kosong, kredensial salah, login berdasarkan peran, input email kosong, deteksi email phishing, deteksi email aman, penyimpanan riwayat, serta pembatasan hak akses. Seluruh skenario menghasilkan status valid. Hasil ini menunjukkan bahwa fungsi utama aplikasi telah berjalan sesuai kebutuhan yang dirancang.

No.	Test Scenario	Result
1	Authentication and role access	Valid
2	Empty or invalid login input	Valid
3	Phishing email classification	Valid
4	Safe email classification	Valid
5	Empty email detection input	Valid
6	Automatic history recording	Valid
7	History access by Staff	Valid
8	History access by Principal	Valid
9	CSV history export	Valid
10	Probability result display	Valid

Table 5. Summary of functional testing

3.6 Discussion

Peningkatan akurasi dari 94,48% menjadi 96,99% menunjukkan bahwa optimasi hyperparameter melalui RSCV memberikan kontribusi positif pada dataset penelitian. Selain akurasi, macro precision, macro recall, dan macro F1-score juga mencapai 97%, sehingga performa kedua kelas menjadi lebih seimbang. Temuan ini mendukung penelitian sebelumnya yang menempatkan Random Forest sebagai salah satu algoritma yang kompetitif untuk klasifikasi phishing [17]–[20], [22].

Jika dibandingkan dengan Dwianti et al. yang memperoleh 98,01% menggunakan TF-IDF dan Random Forest, hasil penelitian ini sedikit lebih rendah. Perbedaan tersebut tidak berarti model penelitian ini lebih buruk secara mutlak karena Dwianti et al. menggunakan dataset publik sebanyak 82.486 email, sedangkan penelitian ini menggunakan 2.989 email nyata dari satu lingkungan sekolah. Penelitian lain juga menunjukkan bahwa kinerja Random Forest dapat berubah menurut pembagian data dan karakteristik dataset [19], [23].

Kontribusi praktis penelitian terletak pada integrasi model ke aplikasi Streamlit. Pengguna tidak perlu menjalankan proses machine learning secara langsung karena sistem menyediakan antarmuka untuk memasukkan email dan memperoleh hasil klasifikasi. Penyajian probabilitas dan XAI juga memberikan informasi tambahan untuk mendukung interpretasi hasil, sejalan dengan perkembangan sistem deteksi phishing yang mulai mempertimbangkan aspek interpretabilitas [21], [24].

Meskipun demikian, terdapat keterbatasan. Dataset berasal dari satu lingkungan sekolah sehingga belum mewakili seluruh variasi email pada institusi lain. Model juga berfokus pada isi teks dan belum menggabungkan header email, metadata pengirim, atau analisis URL secara khusus. Confusion matrix masih menunjukkan satu email phishing yang diklasifikasikan sebagai aman dan 17 email aman yang diklasifikasikan sebagai phishing. Oleh karena itu, sistem sebaiknya digunakan sebagai alat bantu pemeriksaan awal dan perlu diuji pada dataset lintas sekolah serta periode waktu yang berbeda sebelum diterapkan secara lebih luas.

3.8 Limitations and Future Work

Penelitian lanjutan juga dapat mempertimbangkan perbandingan dengan algoritma lain seperti Support Vector Machine, Logistic Regression, Gradient Boosting, atau model bahasa modern. Eksperimen tambahan dapat dilakukan dengan variasi ukuran TF-IDF, strategi balancing, dan teknik validasi yang berbeda. Pengujian lintas dataset juga diperlukan agar dapat diketahui apakah peningkatan performa RSCV tetap konsisten pada data yang memiliki karakteristik berbeda. Dengan pengembangan tersebut, sistem diharapkan tidak hanya memiliki performa klasifikasi yang baik, tetapi juga lebih adaptif terhadap perubahan pola serangan phishing.

3.8 Limitations and Future Work

Keterbatasan utama penelitian terdapat pada sumber dataset yang berasal dari satu lingkungan sekolah dan periode data tertentu. Variasi email dari sekolah lain, perguruan tinggi, organisasi, atau bahasa yang berbeda belum diuji. Selain itu, penelitian menggunakan klasifikasi berbasis teks sehingga informasi teknis seperti reputasi domain, struktur URL, metadata header, attachment, dan hubungan antarentitas belum dimanfaatkan sebagai fitur. Penelitian lanjutan dapat menggabungkan fitur tekstual dengan fitur URL dan metadata serta menguji model pada dataset eksternal untuk mengetahui kemampuan generalisasi.

3.7 Practical Implications

Dari sisi penerapan, sistem dapat digunakan sebagai lapisan pemeriksaan awal sebelum email diproses melalui prosedur keamanan yang lebih lengkap. Ketika sebuah pesan memperoleh prediksi phishing dengan probabilitas tinggi, pengguna dapat melakukan pemeriksaan tambahan terhadap pengirim, tautan, lampiran, dan konteks pesan. Sebaliknya, hasil Safe Email tetap tidak berarti bahwa pesan pasti aman karena model bekerja berdasarkan pola yang dipelajari dari dataset. Posisi sistem sebagai alat bantu awal perlu dipertahankan agar pengguna tidak menganggap hasil klasifikasi sebagai jaminan keamanan mutlak.

3.7 Practical Implications

Jika dibandingkan dengan penelitian terdahulu, hasil 96,99% menunjukkan bahwa pendekatan TF-IDF dan Random Forest tetap kompetitif untuk klasifikasi email phishing. Perbedaan nilai akurasi antarpenelitian dapat dipengaruhi oleh karakteristik dataset, bahasa email, jumlah data, proses preprocessing, pembagian data, serta konfigurasi model. Oleh sebab itu, nilai akurasi tidak dapat dibandingkan secara mutlak tanpa memperhatikan kondisi eksperimen. Kontribusi penelitian ini lebih diarahkan pada pengujian optimasi RSCV pada data email dari lingkungan sekolah dan integrasinya ke aplikasi yang dapat digunakan secara langsung.

3.5 Functional Testing

Black Box Testing terhadap sepuluh skenario menunjukkan bahwa fungsi utama aplikasi dapat dijalankan sesuai kebutuhan pengujian. Skenario mencakup autentikasi, validasi input, pemilihan peran, proses deteksi email, tampilan hasil, penyimpanan riwayat, dan pembatasan akses. Seluruh skenario memperoleh hasil Valid. Meskipun demikian, hasil pengujian fungsional tidak dapat digunakan untuk menyatakan bahwa sistem bebas dari kesalahan keamanan. Pengujian keamanan, pengujian beban, dan validasi model pada data eksternal masih diperlukan sebagai tahap pengembangan berikutnya.

3.4 System Implementation

Riwayat pengujian memberikan manfaat dalam proses pemantauan penggunaan sistem. Data yang disimpan mencakup waktu pengujian, aktor, teks email, hasil prediksi, dan probabilitas. Pencatatan tersebut dapat digunakan untuk melihat pola penggunaan dan membantu evaluasi ketika ditemukan hasil klasifikasi yang perlu diperiksa kembali. Pada sisi hak akses, peran Staf TU dan Principal memiliki fungsi yang berbeda. Pembatasan tersebut penting agar data hasil deteksi dan informasi administratif tidak dapat diakses oleh pengguna yang tidak memiliki kewenangan.

3.4 System Implementation

Fitur XAI pada aplikasi digunakan sebagai informasi pendukung untuk membantu pengguna memahami kata atau karakteristik teks yang berkontribusi terhadap hasil klasifikasi. Fitur ini tidak dimaksudkan sebagai pengganti keputusan keamanan, tetapi sebagai alat bantu interpretasi. Penyajian informasi pengaruh kata dapat membantu pengguna memeriksa kembali isi email ketika model memberikan hasil phishing. Dengan demikian, sistem tidak hanya menghasilkan label akhir, tetapi juga memberikan konteks yang dapat digunakan dalam proses pemeriksaan manual.

3.4 System Implementation

Implementasi aplikasi dilakukan menggunakan Streamlit agar model dapat digunakan melalui antarmuka berbasis web. Pengguna dapat memasukkan isi email ke dalam form yang tersedia, kemudian sistem melakukan preprocessing dan menjalankan model untuk menghasilkan prediksi. Hasil klasifikasi ditampilkan bersama probabilitas sehingga pengguna dapat melihat tingkat keyakinan model. Pendekatan ini membuat proses pengujian lebih praktis dibandingkan menjalankan notebook atau skrip machine learning secara manual. Aplikasi juga memisahkan fungsi sesuai peran pengguna sehingga informasi riwayat dan fungsi administratif tidak dapat diakses secara bebas.

3.3 Confusion Matrix Analysis

Analisis confusion matrix memberikan informasi yang lebih rinci mengenai jenis kesalahan yang terjadi. Pada model teroptimasi, 311 Safe Email berhasil dikenali dengan benar dan 269 Phishing Email juga berhasil diklasifikasikan dengan benar. Sebanyak 17 Safe Email masih diprediksi sebagai phishing, sedangkan hanya satu Phishing Email yang diprediksi sebagai Safe Email. Dari sisi keamanan, kesalahan false negative tersebut perlu menjadi perhatian karena pesan phishing yang dianggap aman berpotensi terlewatkan. Walaupun jumlahnya sangat kecil pada data uji, pengujian pada dataset yang lebih besar dan lebih beragam tetap diperlukan sebelum sistem digunakan sebagai mekanisme keamanan utama.

3.2 Model Performance

Perbandingan model memperlihatkan bahwa Random Forest standar menghasilkan akurasi 94,48%, sedangkan Random Forest setelah optimasi RSCV mencapai 96,99%. Selisih 2,51 poin persentase menunjukkan bahwa konfigurasi parameter memberikan pengaruh terhadap kemampuan model dalam memisahkan dua kelas email. Kenaikan tersebut juga terlihat pada precision, recall, dan F1-score yang meningkat hingga sekitar 97%. Dengan demikian, proses optimasi tidak hanya meningkatkan satu metrik, tetapi memberikan perbaikan yang relatif konsisten pada beberapa ukuran performa.

3.1 Dataset Characteristics

Hasil distribusi dataset menunjukkan bahwa Safe Email memiliki jumlah data yang sedikit lebih besar daripada Phishing Email. Selisih tersebut tidak terlalu jauh sehingga model tetap dapat mempelajari kedua kelas tanpa ketimpangan yang ekstrem. Namun, distribusi yang tidak sepenuhnya seimbang tetap perlu diperhatikan ketika membaca hasil evaluasi. Oleh karena itu, penelitian tidak hanya menggunakan accuracy, tetapi juga precision, recall, dan F1-score. Penggunaan beberapa metrik membantu memberikan gambaran yang lebih lengkap mengenai kemampuan model dalam mengenali masing-masing kelas.

2.2 TF-IDF Feature Extraction

Konfigurasi TF-IDF menggunakan maksimal 1.000 fitur dengan kombinasi unigram dan bigram. Unigram merepresentasikan satu kata, sedangkan bigram memungkinkan model menangkap pasangan kata yang muncul bersama. Pengaturan min_df = 5 digunakan untuk mengabaikan istilah yang muncul sangat sedikit, sementara max_df = 0,7 membatasi pengaruh istilah yang muncul pada sebagian besar dokumen. Konfigurasi tersebut dipilih untuk menjaga jumlah fitur tetap terkendali sekaligus mempertahankan informasi tekstual yang relevan bagi proses klasifikasi.

2.1 Data Acquisition and Preprocessing

Pembagian data menjadi 80% data latih dan 20% data uji dilakukan setelah data selesai diproses. Data latih digunakan untuk membangun model dan melakukan pencarian konfigurasi parameter, sedangkan data uji disimpan untuk evaluasi akhir. Penggunaan cross-validation pada proses RSCV membantu mengurangi ketergantungan terhadap satu pembagian data internal. Dalam penelitian ini digunakan lima fold sehingga setiap bagian data latih memperoleh kesempatan untuk menjadi data validasi. Sepuluh iterasi pencarian acak digunakan untuk mencoba beberapa kombinasi hyperparameter yang tersedia.

2.1 Data Acquisition and Preprocessing

Pada tahap preprocessing, pembersihan teks dilakukan untuk mengurangi karakter atau pola yang tidak diperlukan dalam proses klasifikasi. Case folding digunakan untuk menyeragamkan bentuk huruf, sedangkan alamat email dan URL dimasking agar model tidak terlalu bergantung pada identitas atau alamat tertentu. Stopword removal digunakan untuk mengurangi kata umum yang kurang memberikan informasi pembeda, kemudian stemming dilakukan untuk mengembalikan kata ke bentuk dasarnya. Tahap ini penting karena kualitas fitur yang masuk ke model sangat dipengaruhi oleh kualitas teks yang telah dipersiapkan.

2. RESEARCH METHOD

Secara umum, alur penelitian dimulai dari pengumpulan data email, pemeriksaan kualitas data, preprocessing teks, pembentukan fitur TF-IDF, pembagian data latih dan data uji, pelatihan Random Forest, optimasi hyperparameter dengan RSCV, evaluasi hasil, dan implementasi model terbaik ke aplikasi. Setiap tahap dibuat

berurutan untuk menjaga konsistensi proses eksperimen. Data uji tidak digunakan dalam proses pencarian parameter sehingga pengukuran akhir dapat memberikan gambaran yang lebih objektif terhadap kemampuan model pada data yang tidak digunakan saat pelatihan.

4. CONCLUSION

Penelitian ini menghasilkan sistem deteksi email phishing berbasis machine learning dengan Random Forest yang dioptimasi menggunakan Random Search Cross-Validation. Dari 2.989 email hasil preprocessing, model Random Forest standar memperoleh akurasi 94,48%, sedangkan model hasil optimasi mencapai 96,99%, atau meningkat 2,51 poin persentase. Model optimasi juga memperoleh macro precision, macro recall, dan macro F1-score sebesar 97%. Hasil confusion matrix menunjukkan bahwa model mampu mengenali seluruh email phishing pada data uji, meskipun masih terdapat satu false negative dan 17 false positive. Model terbaik berhasil diimplementasikan ke dalam aplikasi Streamlit yang menyediakan deteksi email, probabilitas hasil, analisis XAI, riwayat pengujian, dan pengaturan hak akses. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam, menggabungkan fitur header, URL, dan metadata email, serta melakukan pengujian lintas institusi agar kemampuan generalisasi model dapat dievaluasi lebih kuat.

REFERENCES

[1] W. Hidayat M., N. Musdira, N. Rasyid, M. K. S., and M. Juharman, "Analisis ancaman terhadap keamanan data pribadi pada email," J. Pendidikan Terapan, vol. 1, pp. 7–12, 2023, doi: 10.61255/jupiter.v1i2.73.

[2] R. Aditya, A. Ramdhani, D. D. Nugraha, and M. F. Ari, "Analisis masalah serangan phishing pada penggunaan email," 2025.

[3] K. Subandi, D. M. Ahmad, V. I. Sugara, and A. S. Aryani, "Awareness cybercrime phishing email menggunakan Secure Socket Layer sebagai interpretasi information security," vol. 9, pp. 56–66, 2024, doi: 10.33751/komputasi.v19i2.5260.

[4] Y. R. Purnamadewi and A. Zahra, "Enhancing detection of zero-day phishing email attacks in the Indonesian language using deep learning algorithms," Bull. Electr. Eng. Informatics, vol. 14, no. 1, pp. 505–512, 2025, doi: 10.11591/eei.v14i1.8759.

[5] C. Herdian, M. Quinn, D. Revaldo, and S. Margareta, "Perbandingan algoritma Naive Bayes di dalam Scikit-Learn Python Library dengan murni algoritma Naive Bayes: Studi kasus klasifikasi email berbahaya," Techno Xplore, vol. 9, no. 1, pp. 1–10, 2024.

[6] R. Nurcahyo, F. Tanjung, and S. Rahman, "Meningkatkan deteksi email phishing melalui pendekatan SVM yang dioptimalkan NLP," 2025.

[7] U. Kurniati, M. A. S. Arifin, and N. K. Daulay, "Penerapan Random Forest dan Naive Bayes untuk deteksi phishing dan serangan social engineering pada email," pp. 891–898, 2025.

[8] D. Septiani and I. Isabela, "Analisis Term Frequency Inverse Document Frequency (TF-IDF) dalam temu kembali informasi pada dokumen teks," SINTESIA J. Sist. dan Teknol. Inf. Indones., vol. 1, no. 2, pp. 81–88, 2022.

[9] I. R. Dwianti, D. R. Cahyani, T. Khawa, A. Rahman, and M. Faisal, "Phishing email classification using TF-IDF method and Random Forest algorithm," vol. 3, no. 2, pp. 125–135, 2025, doi: 10.25181/rt.v3i2.4310.

[10] M. A. Nivedha and S. Raja, "Detection of email spam using Natural Language Processing based Random Forest approach," Int. J. Comput. Sci. Mob. Comput., vol. 11, no. 2, pp. 7–22, 2022, doi: 10.47760/ijcsmc.2022.v11i02.002.

[11] M. Adipa, A. T. Zy, and M. M. Effendi, "Klasifikasi email phishing menggunakan algoritma K-Nearest Neighbor," J. RESTIKOM, vol. 5, no. 2, pp. 148–157, 2023, doi: 10.52005/restikom.v5i2.152.

[12] M. H. Z. Ubaidillah and Z. Fatah, "Klasifikasi email spam menggunakan algoritma Naive Bayes," vol. 2, pp. 129–132, 2024.

[13] M. Fajri and A. Primajaya, "Komparasi teknik hyperparameter optimization pada SVM untuk permasalahan klasifikasi dengan menggunakan Grid Search dan Random Search," J. Appl. Informatics Comput., vol. 7, no. 1, pp. 14–19, 2023, doi: 10.30871/jaic.v7i1.5004.

[14] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk mengevaluasi kinerja mahasiswa," J. Algoritm., vol. 21, no. 1, pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.

[15] R. Tuntun, K. Kusrini, and K. Kusnawi, "Analisis perbandingan kinerja algoritma klasifikasi dengan menggunakan metode K-Fold Cross Validation," J. Media Inform. Budidarma, vol. 6, no. 4, p. 2111, 2022, doi: 10.30865/mib.v6i4.4681.

[16] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi potensial gempa bumi Indonesia menggunakan metode Random Forest dan feature selection," IDEALIS, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealis.v6i2.3036.

[17] A. A. Tawil, L. Almazaydeh, D. Qawasmeh, B. Qawasmeh, M. Alshinwan, and K. Elleithy, "Comparative analysis of machine learning algorithms for email phishing detection using TF-IDF, Word2Vec, and BERT," Comput. Mater. Contin., vol. 81, no. 2, pp. 3395–3412, 2024, doi: 10.32604/cmc.2024.057279.

[18] "Machine learning approach for email phishing detection," Procedia Comput. Sci., vol. 251, pp. 746–751, 2024, doi: 10.1016/j.procs.2024.11.179.

[19] C. Umam, L. B. Handoko, and F. O. Isinkaye, "Performance analysis of Support Vector Classification and Random Forest in phishing email classification," Sci. J. Informatics, vol. 11, no. 2, pp. 367–374, 2024, doi: 10.15294/sji.v11i2.3301.

[20] H. Fares, N. Mouakkal, Y. Baddi, and N. Hajraoui, "Robust email phishing detection using machine learning and deep learning approach," Int. J. Commun. Netw. Inf. Secur., vol. 16, no. 3, 2024, doi: 10.17762/ijcnis.v16i3.6700.

[21] R. Jayaprakash, K. Natarajan, J. A. Daniel, C. V. Chinnappan, J. Giri, H. Qin, and S. Mallik, "Heuristic machine learning approaches for identifying phishing threats across web and email platforms," Front. Artif. Intell., vol. 7, 2024, doi: 10.3389/frai.2024.1414122.

[22] A. Al-Subaiey, M. Al-Thani, N. A. Alam, K. F. Antora, A. Khandakar, S. M. A. U. Zaman, "Novel interpretable and robust web-based AI platform for phishing email detection," Comput. Electr. Eng., vol. 120, p. 109625, 2024, doi: 10.1016/j.compeleceng.2024.109625.

[23] R. A. Fahim, M. S. Arman, I. Sultana, N. Tasnim, K. R. Ahmed, and I. Mahmud, "PhishGuard: Leveraging NLP and machine learning for email phishing detection," in Proc. 2024 Int. Conf. Big Data Analytics in Bioinformatics (DABCon), 2024, pp. 1–6, doi: 10.1109/DABCon63472.2024.10919349.

[24] M. A. Mubarok, N. Anwar, A. Ichwani, and R. T. Iswahyudi, "Studi simulasi deteksi email phishing berbasis NLP menggunakan TF-IDF dengan Support Vector Machine dan Random Forest," IKRA-ITH Informatika, vol. 10, no. 3, pp. 38–44, 2026.

[25] D. Sugianto, R. A. Putawa, C. Izumi, and S. A. Ghaffar, "Uncovering the efficiency of phishing detection: An in-depth comparative examination of classification algorithms," Int. J. Appl. Inf. Manag., vol. 4, no. 1, pp. 22–29, 2024, doi: 10.47738/ijaim.v4i1.72.