Jurnal Informatika dan Teknologi Interaktif
Volume 3, Nomor 1, April 2026

DOI : 10.63547/jiite.v3i1.118
e-ISSN : 3048-1023

Feature Fusion TF-IDF dan Analisis URL untuk Deteksi Phishing Lintas Domain dengan LinearSVC

Gemara Aurelya[1], Ridho Rian Sahputra[2], Hanrifki Pratama[3], Nur Rahma Keysha Maharani[4], Indah Adi Setiaputri[5], Mohammad Rizki Dwi Saputra[6]

[1,2,3,4,5]Jurusan Informatika, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Purwokerto, Indonesia
[6]Jurusan Software Engineering, Fakultas Ilmu Komputer, Universitas Insan Pembangunan Indonesia, Bitung, Tangerang, Indonesia
surel:[1]gemaraaurelya@gmail.com,[2]rihdoriansahputra@gmail.com,[3]hanrifki11@gmail.com, [4]keyshaicham@gmail.com, [5]indahadise18@gmail.com, [6]mohr6977@gmail.com

Info Artikel

Sejarah artikel:
Diterima 26-01-2026
Revisi 15-02-2026
Diterima 07-03-2026

Kata kunci:
phishing email
feature fusion
TF-IDF
LinearSVC
evaluasi lintas domain

ABSTRAK

Meningkatnya serangan phishing melalui email menuntut sistem deteksi otomatis yang tidak hanya akurat, tetapi juga mampu beradaptasi terhadap perbedaan karakteristik data antar domain. Meskipun berbagai pendekatan berbasis *machine learning* telah menunjukkan performa tinggi, sebagian besar penelitian masih berfokus pada penggunaan satu jenis fitur dan evaluasi dalam satu domain, sehingga kemampuan generalisasi model pada data nyata yang heterogen belum teruji secara komprehensif. Penelitian ini bertujuan untuk mengatasi keterbatasan tersebut dengan mengusulkan pendekatan deteksi email phishing yang lebih menyeluruh melalui integrasi fitur tekstual dan struktural serta evaluasi lintas domain. Metode yang diusulkan mengombinasikan fitur tekstual dari badan email yang diekstraksi menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF) dan fitur struktural URL dalam skema feature fusion, dengan proses klasifikasi menggunakan algoritma *Linear Support Vector Classifier* (LinearSVC) yang efisien untuk data berdimensi tinggi dan bersifat sparse. Dataset SpamAssassin digunakan sebagai data pelatihan dan evaluasi internal, sedangkan dataset Enron digunakan untuk evaluasi lintas domain guna mengukur ketahanan model terhadap *domain bias*. Hasil evaluasi internal menunjukkan bahwa model mencapai akurasi sebesar 98,84% dengan nilai F1-score makro 97,75%, sementara pada evaluasi lintas domain model mempertahankan kemampuan deteksi email legitimate dengan nilai *recall* sebesar 93,62% meskipun terjadi penurunan kinerja akibat perbedaan distribusi data. Temuan ini menunjukkan bahwa pendekatan *feature fusion* lebih tangguh dibandingkan penggunaan fitur tunggal dan efektif sebagai *baseline* deteksi phishing lintas domain, serta berpotensi dikembangkan lebih lanjut melalui strategi adaptasi domain untuk meningkatkan ketahanan model pada lingkungan nyata.

Penulis yang sesuai:
Gemara Aurelya
Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Purwokerto
Email: gemaraaurelya@gmail.com

1. PENDAHULUAN

Di era digital saat ini, email telah menjadi tulang punggung komunikasi baik dalam konteks personal maupun profesional. Kemudahan, kecepatan, dan jangkauan luas yang ditawarkan oleh email menjadikannya sarana utama pertukaran informasi di berbagai sektor. Namun, kondisi ini juga dimanfaatkan oleh pelaku kejahatan siber untuk melancarkan serangan phishing. Phishing merupakan bentuk penipuan yang bertujuan memperoleh informasi sensitif, seperti nama pengguna, kata sandi, dan detail kartu kredit, dengan menyamar sebagai entitas tepercaya dalam komunikasi elektronik [1]. Berbagai laporan menunjukkan bahwa sekitar 96% serangan phishing dilakukan melalui email, dan lebih dari 75% organisasi di seluruh dunia pernah menjadi target serangan ini [2]. Dampak yang ditimbulkan tidak hanya berupa kerugian finansial, tetapi juga penurunan kepercayaan terhadap sistem dan platform digital yang digunakan secara luas.

Seiring meningkatnya kompleksitas serangan phishing, berbagai pendekatan berbasis machine learning telah dikembangkan untuk mendeteksi email berbahaya secara otomatis. Meskipun demikian, metode-metode yang ada masih menghadapi sejumlah permasalahan mendasar. Salah satu tantangan utama adalah penurunan kinerja model ketika diaplikasikan pada data yang berasal dari domain yang berbeda dengan data pelatihan. Model yang dilatih menggunakan satu dataset tertentu, misalnya dataset URL phishing dari Kaggle, sering kali tidak mampu mempertahankan akurasinya ketika dihadapkan pada data email korporat atau sumber lain dengan karakteristik berbeda. Fenomena ini dikenal sebagai domain bias dan menjadi hambatan signifikan dalam penerapan sistem deteksi phishing di lingkungan nyata yang bersifat dinamis [3], sebagaimana juga ditegaskan dalam penelitian mengenai generalisasi deteksi phishing lintas domain [4]. Selain itu, banyak penelitian cenderung berfokus pada satu jenis fitur saja, baik fitur berbasis konten teks email [[5], [6]] maupun fitur berbasis struktur URL [7], seperti yang ditunjukkan oleh Mahmud dan Wirawan (2025) dalam deteksi phishing berbasis fitur URL menggunakan algoritma klasifikasi [8]. Pendekatan tunggal tersebut mengabaikan kenyataan bahwa serangan phishing modern umumnya bersifat multi-dimensi yang mengombinasikan teknik rekayasa sosial dalam teks email dengan manipulasi URL yang semakin canggih, sehingga mendorong penggunaan pendekatan feature fusion dalam deteksi phishing [9]. Akibatnya, kurangnya integrasi antara analisis konten dan analisis URL berpotensi menciptakan celah dalam sistem pertahanan.

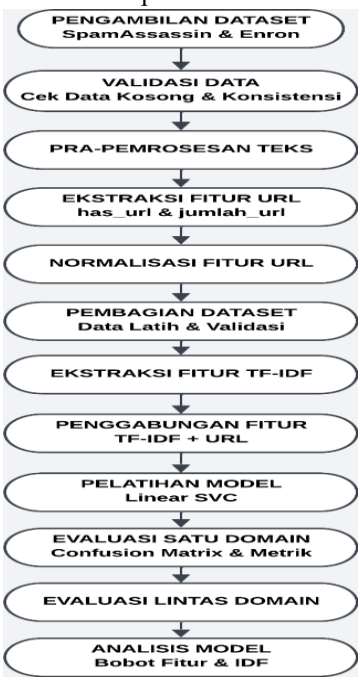
Berbagai penelitian sebelumnya telah menunjukkan bahwa machine learning mampu mencapai performa tinggi dalam deteksi phishing, meskipun dengan ruang lingkup yang masih terbatas. Pradana dan Susanto (2026) mengusulkan model ensemble yang menggabungkan Random Forest, Gradient Boosting, Logistic Regression, dan AdaBoost dengan fokus pada ekstraksi fitur URL, serta berhasil mencapai akurasi sebesar 98,10% pada dataset Kaggle [7]. Namun, penelitian ini hanya mengandalkan fitur URL dan tidak mengevaluasi model pada domain data yang berbeda. Di sisi lain, penelitian yang berfokus pada konten teks email juga menunjukkan hasil yang menjanjikan, baik menggunakan pendekatan tradisional seperti TF-IDF maupun metode representasi modern seperti Word2Vec dan BERT [10]. Utomo et al. (2024) menerapkan Long Short-Term Memory (LSTM) pada dataset SpamAssassin dan memperoleh akurasi 99,35% [11], sementara Rustam et al. (2024) mengombinasikan metode Continuous Bag-of-Words (CBOW) dengan Random Forest dan mencapai akurasi 98,21% pada dataset yang sama [4]. Selain itu, Tanjung dan Rahman (2025) melakukan perbandingan beberapa kernel Support Vector Machine (SVM) dan menemukan bahwa SVM dengan kernel polinomial yang dikombinasikan dengan fitur TF-IDF memberikan akurasi terbaik sebesar 97,85% [12]. Temuan serupa juga dilaporkan oleh Aryanti dan Nabila (2024) yang menerapkan SVM berbasis TF-IDF pada URL phishing dan memperoleh performa klasifikasi yang tinggi [13]. Meskipun hasil yang diperoleh sangat tinggi, sebagian besar penelitian tersebut masih terisolasi pada penggunaan satu jenis fitur dan pengujian dalam satu domain data, sehingga kemampuan generalisasi model belum dievaluasi secara komprehensif. Bahkan ketika menggunakan model hybrid berbasis deep learning dengan optimasi kompleks, tantangan generalisasi lintas domain masih tetap menjadi isu utama [14]. Selain itu, beberapa penelitian terbaru telah mengusulkan integrasi representasi kontekstual berbasis deep learning seperti BERT dengan fitur statistik TF-IDF untuk meningkatkan performa deteksi phishing, khususnya dalam skenario kompleks, meskipun pendekatan tersebut masih memiliki tantangan terkait kompleksitas model dan kebutuhan data yang besar [15].

Berdasarkan keterbatasan tersebut, penelitian ini bertujuan untuk menjembatani celah yang ada dengan mengusulkan pendekatan deteksi email phishing yang lebih menyeluruh dan tangguh. Penelitian ini secara khusus bertujuan untuk mengimplementasikan pendekatan feature fusion dengan menggabungkan fitur tekstual dari badan email yang diekstraksi menggunakan metode TF-IDF dan fitur struktural dari URL yang terkandung di dalam email.

Proses klasifikasi dilakukan menggunakan algoritma Linear Support Vector Classifier (LinearSVC), yang dikenal efisien dalam menangani data berdimensi tinggi seperti data teks. Selain itu, evaluasi kinerja model tidak hanya dilakukan pada data dari domain yang sama (internal evaluation), tetapi juga pada data dari domain yang berbeda (cross-domain evaluation) guna mengukur ketahanan model terhadap domain bias. Dengan pendekatan tersebut, penelitian ini diharapkan dapat menghasilkan model deteksi phishing yang tidak hanya memiliki tingkat akurasi tinggi, tetapi juga mampu beradaptasi dan andal ketika diterapkan pada lingkungan nyata dengan variasi data email yang beragam.

2. METODE

Penelitian ini mengikuti alur kerja sistematis yang dirancang untuk membangun dan mengevaluasi model deteksi phishing secara komprehensif. Metodologi ini mencakup beberapa tahapan utama, mulai dari pengumpulan data hingga analisis model, seperti yang diilustrasikan pada Gambar 1.



Gambar 1. Desain Sistem

Alur penelitian ini dimulai dengan pengumpulan data dari dua sumber berbeda yaitu dataset SpamAssassin dan dataset Enron untuk memungkinkan evaluasi lintas domain. Penggunaan dua dataset dengan karakteristik yang berbeda ini memungkinkan pengujian ketahanan model terhadap data yang belum pernah dilihat sebelumnya. Kemudian pada tahap selanjutnya adalah validasi data untuk memastikan kualitas dan konsistensi. Data kemudian melalui serangkaian langkah pra-pemrosesan, termasuk pembersihan teks dan ekstraksi fitur dari konten email dan URL. Fitur-fitur ini kemudian digabungkan (*feature fusion*) dan digunakan untuk melatih model klasifikasi LinearSVC. Akhirnya, model dievaluasi secara internal pada domain yang sama dan secara eksternal pada domain yang berbeda untuk menguji generalisasinya, diikuti dengan analisis fitur untuk memahami faktor-faktor yang paling berpengaruh dalam deteksi.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari setiap tahapan dalam metodologi penelitian, mulai dari pra-pemrosesan data hingga evaluasi akhir model.

3.1. Dataset

Penelitian ini menggunakan dua dataset email dari domain berbeda, yaitu SpamAssassin Public Corpus dan Enron Email Dataset. Dari kedua dataset tersebut, masing-masing diekstraksi isi teks dan URL yang terdapat di dalamnya. Dari dataset SpamAssassin Public Corpus diperoleh data sebanyak 3.052 email. Sedangkan, jumlah data

pada dataset Enron adalah 517.401 email. Penggunaan dua dataset dengan karakteristik berbeda bertujuan untuk menguji kemampuan generalisasi model dalam mendeteksi email phishing lintas domain.

3.2. Pemeriksaan *Missing Value* dan Pembersihan Data

Pemeriksaan *missing value* dilakukan pada semua kolom dalam dataset SpamAssassin dan Enron. Hasil evaluasi menunjukkan bahwa tidak ada nilai kosong (NaN) yang ditemukan pada kolom email_text, url, maupun label. Namun, ada beberapa data yang memiliki teks atau URL kosong. Dalam dataset SpamAssassin, ditemukan 27 email tanpa teks dan 578 email tanpa URL, sedangkan dalam dataset Enron terdapat 439.474 email tanpa URL. Hal ini bisa terjadi karena tidak semua email menyertakan tautan. Pada tahap pembersihan data, hanya email dalam dataset SpamAssassin yang memiliki email_text kosong yang dihapus, karena teks email adalah fitur utama dalam proses ekstraksi fitur. Setelah proses ini, jumlah data SpamAssassin menjadi 3.025 email, sementara dataset Enron tetap berjumlah 517.401 email.

Tabel 1. Perbandingan Jumlah Data Sebelum dan Sesudah Pembersihan

	Sebelum pembersihan	Sesudah pembersihan
SpamAssassin	3.052	3.025
Enron	517.401	517.401

Tingkat data hilang minimal menunjukkan kualitas dataset baik, sementara perbedaan skala 172 kali lipat dan domain (spam publik vs email korporasi) menciptakan kondisi domain shift ideal untuk evaluasi generalisasi model.

3.3. *Preprocessing* Teks Email

Preprocessing teks email dilakukan melalui serangkaian langkah normalisasi standar untuk menghilangkan noise dan menyederhanakan representasi semantik. Tahapan ini mencakup konversi ke huruf kecil, penghapusan URL, alamat email, angka, serta karakter non-alfabetik menggunakan pola regular expression, diikuti *tokenisasi*, eliminasi *stopwords* bahasa Inggris, dan *stemming* berbasis algoritma Porter. Hasil *preprocessing* menghasilkan teks bersih yang lebih ringkas, dengan pengurangan panjang signifikan sambil mempertahankan esensi konten untuk ekstraksi fitur lanjutan.

Tabel 2. Hasil *Preprocessing* Dataset SpamAssassin

	email_text	clean_text
0	URL: http://www.newsisfree.com/click/-0,859765...	url date last person leav tori parti pleas tur...
1	You can also duplicate thiswith\n\nMsgChange -...	also duplic thiswith msgchang noshow tcl promp...
2	\nSome interesting quotes...\n\nhttp://www.pos...	interest quot thoma jefferson examin known sup...
3	URL: http://www.newsisfree.com/click/-1,840674...	url date img reuter
4	Is there any way to customise the folder table...	way customis folder tabl content specif folder...

Tabel 3. Hasil *Preprocessing* Dataset Enron

	email_text	clean_text
0	To find out about accessories, useful add-on s...	find accessori use add softwar replac part ple...
1	Outside the U.S., please call one of the Tech...	outsid pleas call one tech support phone numbe...
2	Sally -\n\nAre you available for this meeting?...	salli avail meet ask john come also least peop...
3	I would like to continue to receive an email i...	would like continu receiv email morn violat ma...
4	Errol/Kam\n\nFor the meetings that I have sche...	errol kam meet schedul possibl copi dpr produc...

3.4. Ekstrasi Fitur URL

Tabel 4. Fitur URL Pada Sampel Awal Dataset SpamAssassin

	has_url	url_count
0	1	7
1	1	2
2	1	1
3	1	2
4	1	1

Tabel 5. Fitur URL Pada Sampel Awal Dataset Enron

	has_url	url_count
0	1	1

1	1	1
2	0	0
3	0	0
4	0	0

Berdasarkan lima sampel awal dataset SpamAssassin, seluruh email mengandung URL dengan jumlah tautan yang bervariasi antara 1 hingga 7 URL per email dan rata-rata sebesar 2,6 URL, menunjukkan bahwa penggunaan URL merupakan karakteristik umum pada email spam. Sebaliknya, pada dataset Enron, URL hanya ditemukan pada sebagian sampel dengan jumlah yang lebih rendah, sementara email lainnya tidak mengandung URL sama sekali. Perbedaan pola ini mengindikasikan bahwa keberadaan dan jumlah URL berpotensi menjadi fitur diskriminatif dalam klasifikasi email spam dan non-spam.

3.5. Normalisasi Fitur URL

Normalisasi fitur has_url dan url_count dilakukan menggunakan StandardScaler untuk mentransformasikan data ke dalam skala standar dengan mean nol dan deviasi standar satu.

Tabel 6. Hasil Normalisasi Fitur URL Dataset SpamAssassin

	has_url_scaled	url_count_scaled
0	0.471928	2.043729
1	0.471928	0.046223
2	0.471928	-0.353278
3	0.471928	0.046223
4	0.471928	-0.353278

Berdasarkan lima sampel awal pada dataset SpamAssassin, seluruh nilai has_url_scaled bernilai positif dan sama, yang menunjukkan bahwa semua email pada sampel ini mengandung URL. Sebaliknya, nilai url_count_scaled bervariasi dari negatif hingga positif, menandakan perbedaan jumlah URL relatif terhadap rata-rata dataset, sehingga jumlah URL menjadi faktor pembeda yang lebih informatif.

Tabel 7. Hasil Normalisasai Fitur URL Dataset Enron

	has_url_scaled	url_count_scaled
0	0.471928	-0.353278
1	0.471928	-0.353278
2	-2.118966	-0.752779
3	-2.118966	-0.752779
4	-2.118966	-0.752779

Pada dataset Enron, nilai has_url_scaled membentuk dua kelompok yang kontras, merepresentasikan email yang mengandung URL dan yang tidak, sesuai dengan karakteristik biner fitur tersebut. Sementara itu, url_count_scaled didominasi oleh nilai negatif, yang mengindikasikan bahwa sebagian besar email memiliki jumlah URL di bawah rata-rata. Pola ini menunjukkan bahwa keberadaan URL merupakan indikator yang lebih kuat dibandingkan jumlah URL dalam mendukung proses klasifikasi email phishing dan non-phishing.

3.6. *Splitting* Dataset

Dataset SpamAssassin dibagi menjadi dua himpunan utama untuk mendukung proses pelatihan dan evaluasi model klasifikasi spam. Teks yang telah melalui preprocessing (clean_text) dijadikan matriks fitur X_{text}, sementara label biner spam (1) dan non-spam (0) ditetapkan sebagai y. Pembagian dilakukan dengan proporsi 80% untuk pelatihan (X_{train_text}, y_{train}) dan 20% untuk validasi (X_{val_text}, y_{val}) menggunakan fungsi train_test_split dari pustaka scikit-learn.

Tabel 8. Spesifikasi Pembagian Dataset Pelatihan dan Validasi

Himpunan	Proporsi	Tujuan
X_{train_text}, y_{train}	80%	Optimasi parameter model
X_{val_text}, y_{val}	20%	Evaluasi generalisasi

3.7. Ekstrasi TF IDF dan *Feature Fusion* Representasi *Hybrid*

Ekstraksi fitur teks dilakukan dengan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk merepresentasikan konten email ke dalam bentuk vektor numerik berdimensi tinggi. Proses vektorisasi diterapkan pada teks hasil praproses menggunakan TfidfVectorizer dengan batasan 5.000 fitur paling representatif. Selanjutnya, vektor TF-IDF digabungkan secara horizontal (horizontal stacking) dengan dua fitur URL yang telah dinormalisasi, yaitu has_url_scaled dan url_count_scaled, sehingga membentuk representasi fitur hybrid berdimensi 5.002.

Tabel 9. Dimensi Matriks *Feature Fusion*

Himpunan data	Dimensi Matriks	Jumlah sample
Pelatihan	(2420, 5002)	2.420
Validasi	(605, 5002)	605
Enron	(517401, 5002)	517.401

Pada Tabel 9, proses *feature fusion* menghasilkan matriks dengan dimensi yang konsisten pada seluruh himpunan data, yakni (2420, 5002) untuk data pelatihan, (605, 5002) untuk data validasi, dan (517401, 5002) untuk dataset Enron. Konsistensi dimensi ini memastikan kompatibilitas representasi fitur dalam proses pelatihan dan evaluasi model lintas dataset.

Tabel 10. Sampel Awal *Feature Fusion*

	aaron	abandon	abc	abdc	abidjan	...	has_url_scaled	url_count_scaled
0	0.0	0.0	0.0	0.0	0.0	...	0.471405	-0.347687
1	0.0	0.0	0.0	0.0	0.0	...	0.471405	0.808615
2	0.0	0.0	0.0	0.0	0.0	...	0.471405	0.808615
3	0.0	0.0	0.0	0.0	0.0	...	0.471405	0.808615
4	0.0	0.0	0.0	0.0	0.0	...	0.471405	-0.347687

Pada Tabel 10, terlihat sebagian besar fitur TF-IDF bernilai nol akibat sifat sparsitas vektor teks, sementara fitur URL menunjukkan variasi nilai antar sampel. Penggabungan fitur leksikal dan struktural ini menghasilkan representasi hybrid yang mengintegrasikan informasi semantik dan karakteristik URL secara simultan, sehingga menyediakan dasar yang stabil untuk tahap klasifikasi dan evaluasi generalisasi model lintas domain SpamAssassin–Enron.

3.8. Pembobotan IDF

Analisis bobot Inverse Document Frequency (IDF) dilakukan untuk mengidentifikasi term dengan tingkat kelangkaan tertinggi dalam korpus SpamAssassin, yang berpotensi memiliki daya diskriminasi tinggi dalam membedakan email spam dan non-spam. Nilai IDF diurutkan secara menurun untuk menampilkan lima term dengan skor tertinggi dari total 5.000 fitur TF-IDF yang digunakan.

Tabel 11. *Term* Dengan Nilai IDF Tertinggi

term	idf
zowi	8.098789
abmv	8.098789
zwluyi	8.098789
aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa...	8.098789
billington	8.098789

Berdasarkan Tabel 11, *term "zowi"*, *"abmv"*, *"zwluyi"*, rangkaian karakter panjang, dan *"billington"* memiliki nilai IDF tertinggi sebesar 8.098789, yang menunjukkan bahwa term-term tersebut muncul pada jumlah dokumen yang sangat terbatas dalam korpus. Kelangkaan ini mengindikasikan potensi term sebagai fitur leksikal yang informatif, khususnya apabila kemunculannya berkorelasi dengan kelas tertentu, sehingga dapat meningkatkan kemampuan diskriminatif model dalam mendeteksi email spam dan non-spam.

3.9. Pelatihan dan Evaluasi Model Linear SVC

Model Linear *Support Vector Classifier* (LinearSVC) dengan parameter random_state = 42 dilatih menggunakan matriks fitur *hybrid* yang mengombinasikan representasi leksikal TF-IDF dan fitur struktural URL pada dataset SpamAssassin. Evaluasi kinerja model dilakukan menggunakan himpunan validasi independen dengan metrik klasifikasi standar, meliputi precision, recall, F1-score, dan accuracy.

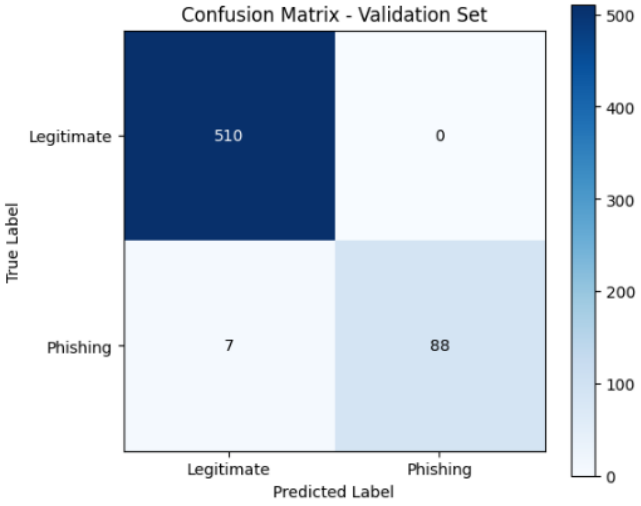
Tabel 12. Matriks Evaluasi Model Linear SVC

	precision	recall	f1-score	Support
0	0.9865	1.0000	0.9932	510
1	1.0000	0.9263	0.9617	95
acccuracy			0.9884	605
Macro avg	0.9932	0.9632	0.9775	605

Hasil evaluasi menunjukkan akurasi model sebesar 98,84%. Pada kelas non-spam (kelas 0), model mencapai nilai precision 98,65% dan recall 100%, menunjukkan kemampuan deteksi email legitimate yang sangat tinggi. Sementara itu, pada kelas spam (kelas 1), nilai precision dan recall masing-masing sebesar 100% dan 92,63%, yang mengindikasikan performa deteksi spam yang baik meskipun terdapat ketidakseimbangan kelas. Nilai F1-score makro sebesar 97,75% menegaskan keseimbangan kinerja model pada kedua kelas.

Hasil ini menunjukkan bahwa representasi fitur *hybrid* efektif dalam menangkap karakteristik leksikal dan struktural email, serta menegaskan LinearSVC sebagai *baseline* yang kuat untuk klasifikasi spam berbasis fitur teks berdimensi tinggi dan bersifat *sparse*. Berdasarkan performa tersebut, model selanjutnya diuji pada dataset Enron berskala besar untuk mengevaluasi kemampuan generalisasi lintas domain.

3.10. Confusion Matrix Evaluasi Validasi



Gambar 2. Confusion Matrix

Berdasarkan confusion matrix pada data validasi, model LinearSVC berhasil mengklasifikasikan seluruh 510 email legitimate dengan benar (true negative) tanpa menghasilkan false positive, yang menunjukkan tingkat kesalahan yang sangat rendah dalam mengidentifikasi email non-phishing. Pada kelas phishing, model berhasil mendeteksi 88 email phishing secara benar (true positive), namun masih terdapat 7 email phishing yang salah diklasifikasikan sebagai legitimate (false negative). Pola ini mengindikasikan bahwa model memiliki kecenderungan lebih konservatif dalam memprediksi kelas phishing, sehingga meskipun tingkat precision tinggi, masih terdapat sejumlah kecil email phishing yang lolos dari deteksi. Secara keseluruhan, distribusi kesalahan yang minim ini menegaskan kinerja model yang sangat baik dan konsisten dengan nilai akurasi serta F1-score yang tinggi pada tahap evaluasi.

3.11. Evaluasi Lintas Domain pada Dataset Enron

Model LinearSVC yang dilatih pada dataset SpamAssassin diuji pada dataset Enron untuk mengevaluasi kemampuan generalisasi lintas domain. Prediksi dilakukan terhadap 517.401 email Enron menggunakan representasi fitur hybrid yang sama.

Tabel 13. Matriks Evaluasi Model Linear SVC Lintas Domain

	precision	recall	f1-score	Support
0	1.0000	0.9362	0.9670	517401
1	0.0000	0.0000	0.0000	0
acccuracy			0.9362	517401
Macro avg	0.5000	0.4681	0.4835	517401

Hasil evaluasi menunjukkan bahwa seluruh data Enron berada pada kelas non-spam, sehingga metrik untuk kelas spam tidak terdefinisi. Model mencapai nilai recall sebesar 93,62% pada kelas non-spam, yang menunjukkan bahwa sebagian besar email legitimate berhasil diklasifikasikan dengan benar, meskipun sekitar 6,38% email mengalami salah klasifikasi sebagai spam. Nilai akurasi sebesar 93,62% perlu ditafsirkan secara hati-hati karena dipengaruhi oleh dominasi satu kelas dalam data uji.

Temuan ini mengindikasikan adanya domain shift antara SpamAssassin dan Enron, yang berdampak pada penurunan kinerja model pada skenario lintas domain, serta menegaskan perlunya strategi adaptasi model untuk meningkatkan robustitas pada penerapan dunia nyata.

3.12. Analisis Fitur Kata Diskriminatif Berbasis TF-IDF

Analisis fitur diskriminatif berbasis TF-IDF dilakukan untuk mengidentifikasi kata-kata yang memiliki kontribusi paling signifikan dalam membedakan email phishing dan email legitimate. Hasil analisis menunjukkan bahwa email phishing didominasi oleh kata-kata seperti click, visit, href, free, money, guarantee, dan unsubscribe. Kata-kata tersebut mencerminkan pola bahasa persuasif dan ajakan bertindak yang umum digunakan dalam email phishing untuk memanipulasi penerima agar mengakses tautan atau memberikan informasi sensitif. Selain itu, kemunculan istilah teknis HTML seperti href dan font mengindikasikan penggunaan struktur tautan tersembunyi sebagai sarana penipuan.

Sebaliknya, email legitimate ditandai oleh kata-kata seperti date, wrote, said, server, list, rpm, dan perl yang bersifat informatif dan kontekstual. Kata-kata tersebut banyak ditemukan pada email komunikasi resmi, diskusi teknis, maupun mailing list, sehingga jarang muncul pada email phishing. Perbedaan karakteristik kata pada kedua kelas menunjukkan bahwa bobot TF-IDF mampu merepresentasikan pola linguistik yang khas dan membedakan antara email phishing dan email legitimate secara efektif.

Hasil ini menunjukkan bahwa metode TF-IDF efektif dalam mengekstraksi fitur tekstual yang bersifat diskriminatif, sehingga mendukung kinerja algoritma Support Vector Machine (SVM) dalam melakukan klasifikasi email phishing secara akurat. Perbedaan distribusi bobot kata pada kedua kelas memperkuat bahwa pendekatan berbasis konten teks relevan untuk digunakan dalam deteksi phishing lintas domain.

4. KESIMPULAN

Penelitian ini mengusulkan pendekatan deteksi email phishing berbasis *feature fusion* dengan mengombinasikan fitur tekstual TF-IDF dan fitur struktural URL menggunakan algoritma Linear Support Vector Classifier (LinearSVC). Hasil evaluasi internal pada dataset SpamAssassin menunjukkan bahwa representasi fitur *hybrid* mampu menghasilkan kinerja klasifikasi yang sangat baik, dengan akurasi sebesar 98,84% dan nilai F1-score makro sebesar 97,75%. Temuan ini menunjukkan bahwa integrasi fitur leksikal dan struktural efektif dalam menangkap karakteristik phishing yang bersifat multidimensional dan meningkatkan kemampuan model dalam membedakan email phishing dan legitimate.

Evaluasi lintas domain pada dataset Enron menunjukkan bahwa model masih mampu mengidentifikasi email legitimate dengan baik, ditunjukkan oleh nilai *recall* sebesar 93,62% pada kelas non-spam, meskipun terjadi penurunan kinerja akibat perbedaan distribusi data antar domain. Hasil ini mengindikasikan adanya pengaruh domain shift terhadap generalisasi model, sekaligus menegaskan bahwa pendekatan feature fusion lebih robust dibandingkan penggunaan fitur tunggal. Secara keseluruhan, pendekatan yang diusulkan efektif sebagai baseline deteksi phishing lintas domain dan berpotensi dikembangkan lebih lanjut melalui strategi adaptasi domain untuk meningkatkan ketahanan model pada lingkungan nyata.

REFERENSI

[1] S. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques," *IEEE Access*, vol. 10, pp. 65703–65727, 2022, doi: 10.1109/ACCESS.2022.3183083.

[2] A. Alhogail and A. Alsabih, "Applying machine learning and natural language processing to detect phishing email," *Comput. Secur.*, vol. 110, Nov. 2021, doi: 10.1016/j.cose.2021.102414.

[3] J. Tanimu, S. Shiaeles, and M. Adda, "A Comparative Analysis of Feature Eliminator Methods to Improve Machine Learning Phishing Detection," *Journal of Data Science and Intelligent Systems*, vol. 2, no. 2, pp. 87–99, Apr. 2024, doi: 10.47852/bonviewJDSIS32021736.

[4] F. Rashid, B. Doyle, S. C. Han, and S. Seneviratne, "Phishing URL detection generalisation using Unsupervised Domain Adaptation," *Computer Networks*, vol. 245, May 2024, doi: 10.1016/j.comnet.2024.110398.

[5] M. Rustam, A. Brotokuncoro, and R. Roestam, "Deteksi Email Spam dengan Continuous Bag-Of-Words dan Random Forest," *R2J*, vol. 6, no. 4, 2024, doi: 10.38035/rrj.v6i4.

[6] R. Nurcahyo, F. Tanjung, and S. Rahman, "Meningkatkan Deteksi Email Phising Melalui Pendekatan SVM yang Dioptimalkan NLP Enhancing Phishing Email Detection through NLP-Optimized SVM Approach." [Online]. Available: http://journal.mahesacenter.org/index.php/incoding

[7] A. Pradana and S. Susanto, “Implementasi Model <i>Machine Learning</i> untuk Deteksi <i>Phishing</i> dengan Pendekatan Ekstraksi Fitur yang Dioptimalkan,” *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 1, pp. 27–40, Jan. 2026, doi: 10.35746/jtim.v8i1.881.

[8] 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,” 2024. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[9] A. Aljofey *et al.*, “An effective detection approach for phishing websites using URL and HTML features,” *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10841-5.

[10] A. Al 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,” *Computers, Materials and Continua*, vol. 81, no. 2, pp. 3395–3412, 2024, doi: 10.32604/cmc.2024.057279.

[11] M. Wahyu Sampurno Utomo, H. Wisnu Murti, A. Widya Indah Sujatmoko, and A. Puspita Sari, “DETEKSI SPAM EMAIL MENGGUNAKAN METODE LSTM (LONG SHORT TERM MEMORY),” 2024.

[12] R. Nurcahyo, F. Tanjung, and S. Rahman, “Meningkatkan Deteksi Email Phising Melalui Pendekatan SVM yang Dioptimalkan NLP Enhancing Phishing Email Detection through NLP-Optimized SVM Approach.” [Online]. Available: http://journal.mahesacenter.org/index.php/incoding

[13] A. Aryanti *et al.*, “Deteksi URL Phishing Menggunakan Algoritma Support Vector Machine Berbasis Website,” 2020, [Online]. Available: https://s.id/jurnalresistor

[14] S. A. A. A. Alsaidi *et al.*, “HawkPhish-DNN cybersecurity model: adaptive hybrid optimization and deep learning for enhanced multi-objective phishing URL detection,” *International Journal of Information Technology (Singapore)*, vol. 17, no. 7, pp. 3859–3875, Sep. 2025, doi: 10.1007/s41870-025-02597-8.

[15] C. M. Chang, M. N. Al-Andoli, and C. Zheng, “Hybrid Phishing Detection Model: Integrating BERT with TF-IDF for Enhanced Email Security,” *International Journal on Robotics, Automation and Sciences*, vol. 7, no. 3, pp. 43–48, Nov. 2025, doi: 10.33093/ijoras.2025.7.3.6.