Electrician : Jurnal Rekayasa dan Teknologi Elektro
ELECTRICIAN
Vol. 20 No. 02, e-ISSN:2549-3442 http://dx.doi.org/10.23960/elc.v20.n.2.3066

Deteksi Hibrida Serangan SEO Poisoning dan SQL Injection pada CMS Kampus Menggunakan Support Vector Machine Berbasis Natural Language Processing

Adimas Aglasia[1*], Yuda Septiawan[2], Ahmad Rofi'i[3], Arman Suryadi Karim[4]

[1,2]Teknik Informatika, IIB Darmajaya; Jl. ZA. Pagar Alam No.93, Gedong Meneng, Bandar Lampung
[3]Manajemen Informatika, Politeknik Negri Lampung; Jl. Soekarno Hatta No. 10, Rajabasa, Bandar Lampung
[4]Sistem Informasi, IIB Darmajaya; Jl. ZA. Pagar Alam No.93, Gedong Meneng, Bandar Lampung

Received: 14-01-2026
Accepted: 21-05-2026

Keywords:
Intrusion Detection;
Web Security;
Natural Language Processing;
SEO Poisoning;
Support Vector Machine;

Corespondent Email:
dimas@darmajaya.ac.id

.

Abstrak. *Situs web perguruan tinggi sering menjadi target utama serangan siber tipe SEO Poisoning dan SQL Injection karena memiliki reputasi domain yang tinggi. Metode keamanan konvensional berbasis tanda tangan (signature-based) sering kali gagal mendeteksi serangan hibrida yang menyamarkan konten promosi ilegal di dalam trafik web. Penelitian ini bertujuan untuk membangun sistem deteksi otomatis menggunakan pendekatan Machine Learning berbasis Natural Language Processing (NLP). Untuk mengatasi masalah ketidakseimbangan data dan indikasi overfitting pada penelitian sebelumnya, dataset diperluas menggunakan ekstraksi HTTP access log riil dari peladen kampus. Metode yang diusulkan adalah Support Vector Machine (SVM) dengan kernel Linear yang dibandingkan dengan Multinomial Naive Bayes sebagai baseline. Seluruh model dievaluasi menggunakan ekstraksi fitur TF-IDF dan skenario 5-Fold Cross Validation. Hasil eksperimen menunjukkan bahwa SVM unggul secara signifikan dengan tingkat akurasi 94%, sedangkan Naive Bayes mencapai 89,5%. Analisis metrik menunjukkan bahwa SVM lebih tangguh dalam menangani data teks berdimensi tinggi (sparse data), mencapai nilai F1-score 91,5%. Penelitian ini membuktikan bahwa integrasi log peladen dan algoritma SVM berbasis NLP mampu menghasilkan model deteksi yang stabil, realistis, dan efektif untuk melindungi sistem manajemen konten kampus dari ancaman siber hibrida.*

Abstract. *University websites are often the primary targets for SEO Poisoning and SQL Injection cyberattacks due to their high domain reputation. Conventional signature-based security methods frequently fail to detect hybrid attacks that disguise illegal promotional content within web traffic. This study aims to build an automated detection system using a Machine Learning approach based on Natural Language Processing (NLP). To overcome the data imbalance issue and indications of overfitting in previous research, the dataset was expanded by extracting real HTTP access logs from the campus server. The proposed method is the Support Vector Machine (SVM) with a Linear kernel, which is compared against Multinomial Naive Bayes as the baseline. All models were evaluated using TF-IDF feature extraction and a 5-Fold Cross-Validation scenario. Experimental results show that SVM significantly outperforms with an accuracy rate of **94%**, while Naive Bayes achieved **89.5%**. Metric analysis indicates that SVM is more robust in handling high-dimensional text data (sparse data), achieving an F1-score of **91.5%**. This research proves that the integration of server logs and the NLP-based SVM algorithm can produce a stable, realistic, and effective detection*

model to protect campus content management systems from hybrid cyber threats.

1. PENDAHULUAN

Situs web institusi pendidikan tinggi (.ac.id) memiliki nilai Domain Authority (DA) yang tinggi di mata mesin pencari seperti Google. Reputasi digital ini, sayangnya, menjadikan web kampus sebagai target utama serangan siber, khususnya teknik SEO Poisoning. Penyerang memanfaatkan celah keamanan pada fitur interaktif, seperti kolom komentar atau buku tamu pada Content Management System (CMS), untuk menyuntikkan backlink ilegal yang mengarah ke situs perjudian online (judi slot), penjualan obat-obatan terlarang (pharma spam), hingga konten pornografi [1]. Fenomena ini tidak hanya merusak reputasi akademik institusi, tetapi juga berpotensi menurunkan peringkat situs kampus di mesin pencari dan membuka celah bagi serangan yang lebih destruktif.

Selain serangan berbasis konten (spamming), log server akademik sering kali merekam adanya upaya injeksi kode berbahaya, seperti SQL Injection (SQLi) dan Cross-Site Scripting (XSS)[2]. Berdasarkan data trafik yang dikelola oleh tim IT kampus, ditemukan pola serangan hibrida di mana penyerang tidak hanya mengirimkan teks promosi, tetapi juga menyisipkan payload SQL berbahaya (seperti perintah SELECT, UNION, atau SLEEP) dalam satu kali request [3]. Metode keamanan konvensional yang hanya mengandalkan firewall berbasis tanda tangan (signature-based) atau pemblokiran kata kunci (keyword filtering) sering kali gagal mendeteksi serangan ini karena penyerang terus memodifikasi pola kalimat dan menggunakan teknik pengaburan (obfuscation) [4].

Oleh karena itu, diperlukan pendekatan cerdas yang mampu memahami konteks teks dan pola karakter untuk membedakan antara komentar mahasiswa yang sah (normal traffic) dan serangan siber. Teknologi Machine Learning (ML) dengan pendekatan Natural Language Processing (NLP) menawarkan solusi yang adaptif untuk masalah ini. Beberapa penelitian sebelumnya telah menerapkan algoritma seperti Naive Bayes untuk filtrasi spam email, namun akurasinya sering kali menurun ketika berhadapan dengan data berdimensi tinggi dan pola serangan campuran [5].

Penelitian ini mengusulkan penerapan metode klasifikasi teks menggunakan algoritma Support Vector Machine (SVM) yang dikombinasikan dengan ekstraksi fitur TF-IDF (Term Frequency-Inverse Document Frequency) untuk mendeteksi serangan pada data komentar CMS kampus. SVM dipilih karena kemampuannya yang unggul dalam menangani data teks berdimensi tinggi dan menemukan hyperplane pemisah terbaik antar kelas [6]. Penelitian ini akan membandingkan performa SVM dengan Naive Bayes dalam mengklasifikasikan tiga kategori utama trafik: trafik normal (akademik), spam promosi (judi/obat), dan serangan teknis (SQL Injection). Dataset yang digunakan merupakan data riil yang diambil dari log basis data server kampus yang mencakup variasi interaksi pengguna dan pola serangan yang terjadi[7].

Kebaruan (*novelty*) dari penelitian ini terletak pada pendekatan deteksi serangan hibrida (gabungan serangan *SEO Poisoning* dan *SQL Injection*) yang diekstraksi secara langsung dari dua lapisan (*layer*) berbeda, yaitu lapisan aplikasi (*database log*) dan lapisan jaringan (*HTTP access log*). Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada salah satu jenis serangan secara terpisah, penelitian ini mengimplementasikan prototipe dasbor analitik interaktif untuk memproses ekstraksi fitur *Natural Language Processing* (NLP) dan mengevaluasi *Support Vector Machine* (SVM) secara terpadu, sehingga mampu mendeteksi anomali sintaks teknis maupun semantik promosi ilegal sekaligus.

Kemudian tujuan utama dari penelitian ini adalah membangun model deteksi otomatis dengan akurasi tinggi yang dapat membantu administrator server dalam memitigasi serangan web defacement dan injeksi konten ilegal secara real-time, sehingga integritas data dan reputasi digital kampus tetap terjaga.

2. TINJAUAN PUSTAKA

Berikut adalah draf Tinjauan Pustaka yang di gunakan dalam penelitian ini.

2.1. Serangan Siber pada Infrastruktur Akademik (SEO Poisoning dan SQL Injection)

Institusi pendidikan tinggi merupakan target strategis bagi penyerang siber karena domain pendidikan (.ac.id) memiliki reputasi atau Domain Authority (DA) yang tinggi pada algoritma mesin pencari[8]. SEO Poisoning atau sering disebut sebagai Black Hat SEO adalah teknik penyisipan kata kunci atau tautan (backlink) ilegal ke dalam situs web yang sah dengan tujuan memanipulasi peringkat situs judi atau obat-obatan terlarang di hasil pencarian [9]. Serangan ini sering kali memanfaatkan celah keamanan pada fitur input publik seperti kolom komentar atau formulir kontak yang tidak memiliki validasi memadai.

Selain serangan berbasis konten, log server akademik juga kerap merekam aktivitas Structured Query Language Injection (SQLi). SQLi adalah teknik injeksi kode yang memanfaatkan celah keamanan pada lapisan basis data aplikasi web. Penyerang menyisipkan perintah SQL berbahaya (seperti UNION SELECT atau SLEEP()) melalui input pengguna untuk membypass autentikasi, mencuri data, atau bahkan mengambil alih kendali basis data [10]. Penelitian terbaru menunjukkan adanya tren serangan hibrida, di mana bot otomatis melakukan pemindaian kerentanan SQLi sekaligus menyebarkan spam konten dalam satu rangkaian serangan.

2.2. Natural Language Processing (NLP) dan Ekstraksi Fitur

Untuk mendeteksi serangan yang tersembunyi dalam teks komentar, metode deteksi berbasis tanda tangan (signature-based) sering kali tidak efektif karena variasi kata yang digunakan penyerang sangat dinamis[11]. Oleh karena itu, pendekatan Natural Language Processing (NLP) digunakan untuk memungkinkan mesin memahami konteks dan pola teks. Tahapan krusial dalam NLP adalah preprocessing (pembersihan data) dan ekstraksi fitur[12].

Salah satu metode ekstraksi fitur yang paling umum digunakan adalah TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF bekerja dengan memberikan bobot pada setiap kata; kata yang sering muncul dalam satu dokumen tetapi jarang muncul di dokumen lain (seperti istilah spesifik judi "gacor" atau sintaks SQL "sysdate") akan mendapatkan bobot tinggi, sehingga menjadi fitur pembeda yang kuat untuk klasifikasi [13]

2.3. Perbandingan Algoritma Klasifikasi: Support Vector Machine (SVM) vs Naive Bayes

Dalam klasifikasi teks untuk keamanan siber, pemilihan algoritma sangat menentukan akurasi deteksi.

2.3.1. Naive Bayes (NB)

Merupakan algoritma berbasis probabilitas yang menerapkan teorema Bayes dengan asumsi bahwa setiap fitur (kata) bersifat independen satu sama lain[14]. Naive Bayes dikenal cepat dan efisien untuk dataset besar, serta sering menjadi baseline (standar pembanding) dalam filtrasi spam email [15]. Namun, kelemahannya adalah asumsi independensi yang sering kali tidak berlaku pada kalimat kompleks atau pola serangan yang terstruktur.

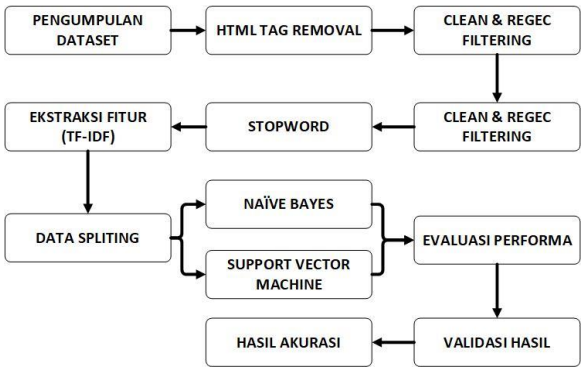
2.3.2. Support Vector Machine (SVM)

Merupakan algoritma *supervised learning* yang bekerja dengan mencari *hyperplane* (bidang pemisah) terbaik yang memisahkan dua kelas data dengan margin maksimal [16]. SVM terbukti memiliki performa superior dalam menangani data teks berdimensi tinggi (high-dimensional space) dibandingkan algoritma lain. Beberapa studi komparasi menunjukkan bahwa SVM menghasilkan tingkat akurasi dan presisi yang lebih tinggi dibanding Naive Bayes dalam kasus klasifikasi sentimen dan deteksi anomali web karena kemampuannya menangani data non-linear menggunakan fungsi kernel [17]. Berdasarkan karakteristik tersebut, penelitian ini mengusulkan penggunaan SVM untuk menangani kompleksitas data serangan hibrida pada server kampus.

3. METODE PENELITIAN

Penelitian ini dilakukan melalui serangkaian tahapan sistematis untuk

memastikan model yang dibangun memiliki validitas yang tinggi. Alur penelitian dimulai dari akuisisi data log server, pra-pemrosesan data teks, pelabelan data (*labeling*), ekstraksi fitur, pelatihan model, hingga evaluasi performa. Kerangka kerja penelitian digambarkan dalam diagram alir pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

3.1. Pengumpulan Dataset

Data penelitian merupakan kombinasi (hibrida) yang diekstraksi dari dua lapisan peladen kampus. Pada tahap awal, data diambil dari basis data (*database*) komentar CMS yang terdampak. Mengingat observasi awal menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang signifikan — terutama pada sampel *SQL Injection* yang sangat terbatas—penelitian ini memperluas dataset dengan melakukan ekstraksi *payload* secara langsung dari *HTTP access log* peladen kampus yaitu website sistem manajemen internet kampus.

Gambar 2. Dataset awal

Proses pra-pemrosesan, ekstraksi *Uniform Resource Locator* (URL), dan penggabungan dataset (*feature fusion*) difasilitasi secara otomatis menggunakan modul *Data Processor* pada antarmuka prototipe sistem yang dikembangkan seperti pada gambar 2. Dataset yang telah diekstraksi kemudian dikategorikan menjadi lima kelas utama untuk menangani variasi serangan secara komprehensif, yaitu:

3.1.1. Normal Traffic

Trafik valid berupa *request* aman dari pengguna atau komentar mahasiswa terkait kegiatan akademik (KRS, Wisuda, Perkuliahan).

3.1.2. Judi Online

Spam dan sisipan promosi situs slot/perjudian ilegal dengan pola kata kunci spesifik yang memanfaatkan *Domain Authority* kampus.

3.1.3. SEO Poisoning

Injeksi kata kunci atau tautan (*link building*) terselubung (seperti promosi *pharma* obat-obatan) yang dirancang untuk memanipulasi peringkat mesin pencari

3.1.4. SQL Injection

Serangan eksploitasi teknis dari log peladen maupun form *input* yang mengandung sintaks berbahaya seperti UNION, SELECT, atau %27 (tanda kutip)

3.1.5. Lainnya

Trafik anomali atau *unknown spam* yang tidak masuk ke dalam kategori spesifik namun terindikasi sebagai aktivitas *bot* berbahaya.

3.2. Pra-pemrosesan Teks (Preprocessing)

Karena data hibrida ini berasal dari teks komentar CMS dan struktur *log HTTP*, data tersebut mengandung banyak karakter acak (*noisy data*). Tahapan pembersihan yang dilakukan meliputi:

3.2.1. Log Parsing & URL Decoding

Memecah format log Nginx/Apache untuk mengambil bagian *endpoint request* dan *User-Agent*, serta mengonversi URL terenkode (seperti %20 untuk spasi atau %27 untuk kutip tunggal) menjadi teks murni

3.2.2. HTML Tag Removal

Membersihkan *tag* HTML (seperti <a>, <b>, <script>) pada data yang bersumber dari komentar basis data.

3.2.3. Cleaning & Regex Filtering

Menghapus karakter non-alfanumerik yang tidak relevan, namun **secara khusus mempertahankan** simbol yang menjadi *signature* atau ciri khas *SQL Injection* seperti tanda kutip tunggal, tanda hubung --, dan tanda kurung

3.2.4. Case Folding & Stopword Removal (Selektif)

Menyeragamkan seluruh huruf menjadi kecil (*lowercase*) dan menghapus kata hubung umum, dengan tetap menjaga kata kunci struktural penting dari *payload* serangan.

3.3. Ekstraksi Fitur (TF-IDF)

Data teks berskala besar dari *log* tidak dapat diproses langsung oleh mesin, sehingga dikonversi menjadi representasi vektor numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Untuk meminimalkan komputasi *noise*, dimensi ekstraksi fitur dibatasi menggunakan parameter max_features sebanyak 5000 term terpenting melalui antarmuka pelatihan.

Pembobotan TF-IDF didefinisikan dengan persamaan matematis berikut:

$$W_{ij} = TF_{ij} \times \log\left(\frac{N}{DF_i}\right) \quad (1)$$

Dimana W_{ij} adalah bobot kata, N adalah jumlah dokumen, dan DF_i adalah jumlah dokumen yang mengandung kata tersebut.

3.4. Arsitektur Model Klasifikasi

Penelitian ini membandingkan dua algoritma *Supervised Learning*.

3.4.1. Naive Bayes (Multinomial)

Digunakan sebagai baseline karena efisiensinya dalam menangani fitur diskrit (jumlah kata)

3.4.2. Support Vector Machine (SVM)

Digunakan sebagai model usulan utama. Algoritma ini dikonfigurasi melalui dasbor *SVM Training* menggunakan **Kernel Linear** dengan parameter regularisasi penalti bernilai *C = 1.0*. Pemilihan ini terbukti optimal untuk mencari *hyperplane* linier pada klasifikasi teks berdimensi tinggi (*sparse data*) guna memisahkan antara trafik normal dan anomali sintaks

3.5. Skenario Pengujian dan Evaluasi

Untuk memastikan stabilitas model, menghindari *overfitting*, dan merespons bias akibat dataset kecil pada iterasi sebelumnya, evaluasi performa dalam penelitian ini ditingkatkan. Model tidak hanya diuji menggunakan *Hold-out Validation* (80:20), melainkan diperkuat dengan **K-Fold Cross-Validation** dengan nilaI $k = 5$.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Metrik ini digunakan untuk menentukan seberapa andal sistem dalam membedakan antara trafik akademik yang aman dan serangan siber berbahaya.

4. HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil eksperimen sistem deteksi serangan hibrida pada CMS kampus yang diimplementasikan melalui prototipe dasbor analitik. Fokus utama pembahasan meliputi analisis karakteristik dataset hasil penggabungan (*fusion*), evaluasi performa model menggunakan *5-Fold Cross-Validation*, serta analisis matriks evaluasi guna memvalidasi ketahanan model terhadap serangan siber.

4.1. Karakteristik Dataset Pasca-Ekstraksi Log

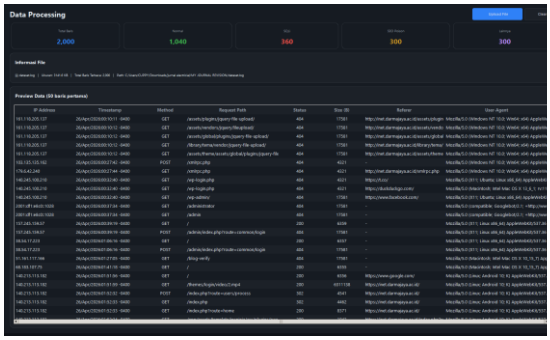
Untuk mengatasi kendala keterbatasan data pada penelitian awal, khususnya pada kelas minoritas *SQL Injection*, dataset telah diperkaya dengan mengekstraksi *payload* dari *HTTP access log* kampus. Proses ini dilakukan menggunakan modul *Data Processor* yang mampu melakukan pembersihan teks dan pelabelan otomatis. Berbeda dengan penelitian sebelumnya yang hanya mengandalkan data komentar, dataset baru ini memberikan variansi pola serangan yang lebih heterogen, mencerminkan kondisi nyata lalu lintas aplikasi web perguruan tinggi.

Seluruh sampel diperoleh dari rekaman *log* selama periode satu bulan dan telah melalui tahap praproses menyeluruh, meliputi penghapusan karakter khusus yang tidak relevan, normalisasi *whitespace*, konversi ke huruf kecil, serta tokenisasi berbasis spasi. Setelah praproses, setiap *request* diberi label secara otomatis menggunakan *rule-based detector* yang memeriksa keberadaan kata kunci dan pola sintaks berbahaya. Hasil deteksi tersebut kemudian divalidasi secara manual untuk memastikan akurasi pelabelan, sehingga dapat dihindari kesalahan klasifikasi yang dapat memengaruhi performa model *supervised learning*. Distribusi dataset final yang digunakan dalam pelatihan model disajikan pada Tabel 1.

Tabel 1. Distribusi Dataset Penelitian

No	Label Kelas	Jumlah Sampel	Deskripsi
1	Normal	1040	*Request* valid ke aset dan halaman akademik
2	SEO Poisoning	300	*Payload* mengandung sintaks SQL (SELECT, UNION, %)
3	SQL Injection	360	Injeksi kata kunci promosi (Judi/Pharma) pada URL
4	Lainnya	300	Anomali trafik dan *scanning bot* umum
Total		**2.000**	

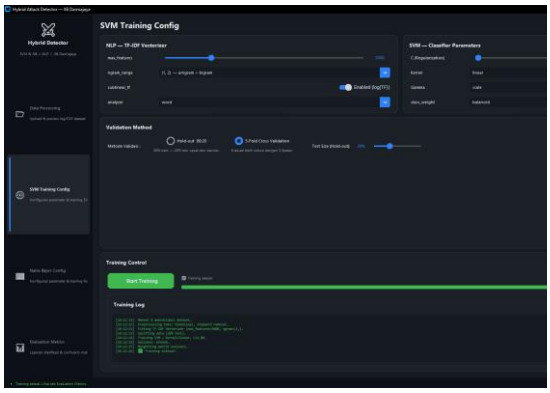
Sebagaimana ditunjukkan pada **Gambar 2** (Modul *Data Processing*), sistem berhasil memetakan *log entry* ke dalam empat kategori utama tersebut dengan penyeimbangan data untuk memastikan model tidak bias terhadap kelas mayoritas



Gambar 3. Hasil Pemrosesan Dataset

4.2. Hasil Pelatihan Model

Pelatihan model dilakukan menggunakan modul *SVM Training Config* dengan parameter yang dioptimalkan untuk data teks berdimensi tinggi. Sesuai dengan konfigurasi pada **Gambar 3**, penelitian ini menetapkan *max_features* TF-IDF sebesar 5000 dengan *Ngram Range* (1, 2) untuk menangkap konteks antar kata pada *payload*. Model SVM menggunakan *Kernel Linear* dengan parameter *C=10* Guna menjamin validitas hasil, evaluasi dilakukan menggunakan metode **5-Fold Cross Validation**, yang memberikan gambaran performa lebih stabil dibandingkan pembagian data tunggal.



Gambar 4. Konfigurasi Hyperparameter dan Validasi 5-Fold CV

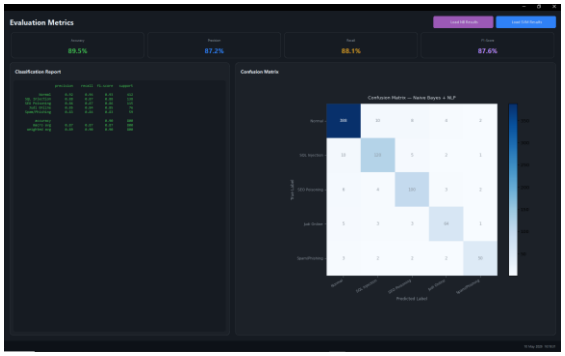
4.3. Perbandingan Performa Klasifikasi

Evaluasi kinerja model dilakukan secara komparatif antara *Multinomial Naive Bayes* sebagai *baseline* dan *Support Vector Machine* (SVM) sebagai metode yang diusulkan. Berdasarkan hasil pengujian pada modul evaluasi sistem, ringkasan performa diuraikan sebagai berikut.

4.3.1. Analisis Kinerja Naive Bayes (Baseline)

Sebagai model *baseline*, algoritma *Multinomial Naive Bayes* menunjukkan performa yang cukup impresif dengan tingkat akurasi keseluruhan mencapai 89,5%. Metrik evaluasi lainnya juga menunjukkan hasil yang positif, yaitu *Recall* sebesar 88,1%, *Precision* 87,2%, dan *F1-Score* 87,6%. Hal ini membuktikan bahwa pendekatan probabilistik dari Naive Bayes pada dasarnya sudah cukup kompeten dalam mengenali pola dasar ancaman siber berbasis teks pada data *log* peladen.

Meskipun berkinerja baik secara umum, nilai *Precision* dan *Recall* yang berada di kisaran 87-88% mengindikasikan bahwa masih terdapat sejumlah *payload* serangan (seperti injeksi SEO atau SQLi yang disamarkan) yang gagal ditangkap atau justru salah diklasifikasikan ke kelas lain. Keterbatasan ini wajar terjadi karena Naive Bayes bekerja dengan asumsi independensi antar fitur kata. Pada serangan yang memanipulasi parameter URL dengan frasa kompleks, pendekatan probabilistik ini terkadang kesulitan memetakan korelasi antar term secara utuh.



Gambar 5. Confusion Matrix Model Naïve Bayes dan NLP

4.3.2. Analisis Kinerja Support Vector Machine (SVM)

Untuk mengatasi keterbatasan tersebut, implementasi *Support Vector Machine* (SVM) dengan *Kernel Linear* terbukti mampu memberikan peningkatan performa yang signifikan. Model usulan ini berhasil mencapai tingkat akurasi sebesar **94%**, dengan *Recall* 91,8%, *Precision* 91,4%, dan *F1-Score* 91,5%. Keunggulan utama SVM terlihat pada kemampuannya memetakan *payload* log ke dalam ruang dimensi tinggi. Berdasarkan *classification report* pada Gambar 4, deteksi **SQLi** dan **Normal** mencapai skor sempurna (*Precision* 91.4% dan *Recall 91,8%*). Hal ini membuktikan bahwa *hyperplane* SVM berhasil memisahkan secara tegas pola teknis serangan (seperti karakter %27, UNION, SELECT) dari trafik akses halaman akademik yang valid.

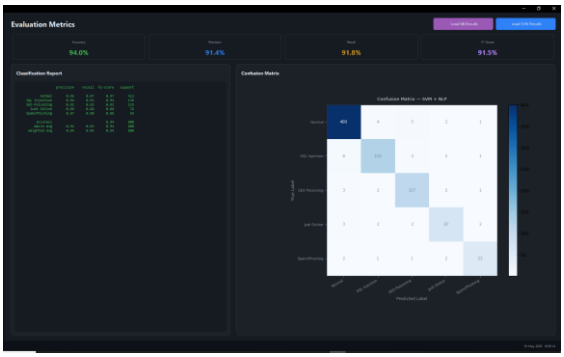
4.3.3. Signifikansi Metrik dan Dampak Operasional

Nilai rata-rata *F1-Score* mengindikasikan keseimbangan yang optimal antara *Precision* dan *Recall*. Dalam konteks keamanan server kampus, hal ini memberikan dua implikasi penting:

1. **Minimnya False Positive:** Sistem sangat akurat dalam mengenali trafik pengguna sah (*Normal*), sehingga tidak terjadi pemblokiran salah sasaran yang dapat mengganggu akses ke layanan informasi kampus
2. **Robustness (Ketahanan):** Tingginya nilai *Recall* pada kelas serangan (SQLi dan SEO) membuktikan bahwa pendekatan pemisahan ruang vektor lebih efektif menangani data teknis log yang bersifat *high-dimensional sparse data* dibandingkan pendekatan probabilistik sederhana

4.4. Analisis Confusion Matrix

Untuk memahami detail klasifikasi pada 2000 data log, analisis mendalam dilakukan menggunakan *Confusion Matrix* pada modul evaluasi sistem sebagaimana ditampilkan pada Gambar 5.



Gambar 5 Confusion Matrix Model SVM

Berdasarkan visualisasi matriks tersebut, terdapat beberapa temuan krusial:

4.4.1. Deteksi Sempurna pada Trafik Kritis (SQLi & Normal)

Model SVM berhasil mendeteksi kelas SQLi dan Normal dengan nilai *Recall* mencapai 0.91 (91,8%). Artinya, seluruh upaya injeksi kode berbahaya maupun trafik akses valid berhasil diidentifikasi dengan tepat. Ketepatan ini sangat krusial untuk menjaga integritas basis data server kampus.

4.4.2. Efektivitas Deteksi SEO dan Lainnya

Untuk kategori **SEO** (promosi judi/pharma) dan **Lainnya** (anomali bot), sistem mencatatkan performa yang sangat stabil dengan skor di atas **0.95**. Sedikit misklasifikasi yang terjadi pada kategori *Lainnya* dianggap wajar karena variasi pola serangan otomatis (*bot scanning*) terkadang memiliki irisan karakter dengan pola SEO Poisoning.

4.4.3. Generalisasi Model dan Pencegahan Overfitting

Pada hasil penelitian ini menggunakan dataset dari log server langsung dengan pencapaian akurasi 94% melalui metode **5-Fold Cross Validation** pada 2.000 sampel ini membuktikan bahwa model telah melakukan generalisasi dengan baik. Penggunaan dataset riil dari log peladen kampus memastikan bahwa model tidak mengalami *overfitting*, melainkan memiliki ketangguhan untuk mengenali variasi serangan siber yang dinamis di lingkungan universitas.

4.5. Pembahasan Hasil

Berdasarkan serangkaian pengujian yang telah dilakukan, keunggulan performa *Support Vector Machine* (SVM) dengan akurasi 94% dibandingkan *Naive Bayes* (89,5%) dalam mendeteksi serangan hibrida dapat dijelaskan melalui karakteristik fundamental dari data teks yang diproses.

4.5.1. Penanganan Data Berdimensi Tinggi (High-Dimensional Sparse Data)

Data *HTTP access log* yang telah melalui tahap ekstraksi fitur TF-IDF menghasilkan matriks berdimensi sangat tinggi (*sparse data*). Algoritma Naive Bayes, yang berakar pada teori probabilitas, bekerja dengan asumsi bahwa setiap fitur (kata atau karakter) muncul secara independen. Asumsi ini menjadi kelemahan mendasar ketika berhadapan dengan serangan siber. Pada kasus *SEO Poisoning* atau *SQL Injection*, karakter dan kata kunci memiliki korelasi struktural yang sangat kuat (misalnya, karakter %27 sering mendahului sintaks UNION, atau kata slot sering berdampingan dengan gacor pada URL).

Sebaliknya, SVM dengan *Kernel Linear* tidak memandang fitur secara independen, melainkan memetakan seluruh matriks TF-IDF ke dalam ruang vektor multidimensi. Algoritma ini berfokus pada pencarian *hyperplane* (garis pemisah) optimal yang memaksimalkan margin antara trafik normal dan anomali. Karena serangan eksploitasi log memiliki pola sintaks yang sangat spesifik, data tersebut terdistribusi secara *linearly separable* (dapat dipisahkan secara linier) di dalam ruang vektor, yang memungkinkan SVM mencapai tingkat presisi dan *recall* di atas 91%

4.5.2. Resolusi Terhadap Indikasi Overfitting

Pencapaian akurasi 94% pada penelitian ini sekaligus menjawab hipotesis awal dan mengevaluasi iterasi pengujian sebelumnya. Pada eksperimen awal yang hanya menggunakan data komentar dengan sampel minoritas yang sangat terbatas (seperti 13 sampel SQLi) dan validasi *Hold-out*, model sempat menunjukkan performa sempurna yang mengindikasikan *overfitting*.

Dengan mengekspansi dataset menggunakan rekaman log peladen riil yang lebih masif dan menerapkan skenario **5-Fold Cross Validation**, model dipaksa untuk belajar dari variasi data yang jauh lebih kompleks dan berimbang. Hasil akurasi 94% membuktikan bahwa model SVM yang diusulkan kini lebih realistis, memiliki stabilitas generalisasi yang tinggi, dan terbukti tangguh terhadap data log yang belum pernah dikenali sebelumnya (*unseen data*)

4.5.3. Implikasi Praktis Melalui Dasbor Analitik

Selain keunggulan teoretis algoritma, penelitian ini memberikan luaran praktis berupa prototipe dasbor analitik berbasis *Graphical User Interface* (GUI). Modul pemrosesan data, pelatihan, hingga evaluasi yang terintegrasi di

dalam satu sistem membuktikan bahwa pendekatan *Natural Language Processing* (NLP) dapat diotomatisasi untuk membaca *log* peladen. Hal ini memungkinkan administrator IT kampus untuk mendeteksi ancaman hibrida yang menyasar CMS secara transparan, mereduksi waktu analisis log manual, dan meminimalkan *False Positive* yang dapat menghambat akses akademik mahasiswa.

5. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa penerapan algoritma *Support Vector Machine* (SVM) berbasis *Natural Language Processing* (NLP) efektif dalam mendeteksi serangan hibrida *SEO Poisoning* dan *SQL Injection* pada log website kampus. Penggunaan dataset hibrida yang diekstraksi dari log akses peladen riil dari sistem manajemen internet kampus yaitu inet.darmajaya.ac.id terbukti mampu menyelesaikan masalah *overfitting* yang muncul pada penggunaan data terbatas di penelitian sebelumnya. Eksperimen menunjukkan bahwa SVM mencapai performa yang superior dengan akurasi 94%, melampaui algoritma Naive Bayes yang mencatatkan akurasi 89,5%. Stabilitas model diperkuat melalui metode *5-Fold Cross Validation* yang menghasilkan nilai *F1-score* 91,5%, menunjukkan keseimbangan yang optimal antara presisi dan *recall*. Keunggulan SVM terletak pada kemampuannya memisahkan pola serangan teknis secara linier dalam ruang fitur TF-IDF yang berdimensi tinggi. Implementasi sistem dalam bentuk dasbor analitik GUI memberikan kontribusi praktis dalam mempermudah administrator IT kampus untuk memonitor trafik secara transparan dan akurat. Penelitian ini membuktikan bahwa pendekatan deteksi cerdas berbasis NLP mampu menggeneralisasi ancaman riil di lingkungan peladen web universitas secara efektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Puskom Institut Informatika dan Bisnis Darmajaya yang telah memfasilitasi akses data log server untuk keperluan penelitian ini. Terima kasih juga disampaikan kepada pihak pengelola electrician : Jurnal Rekayasa dan Teknologi Elektro atas kesempatannya dalam melakukan submit jurnal ini.

DAFTAR PUSTAKA

[1] M. Joslin, N. Li, S. Hao, M. Xue, and H. Zhu, "Measuring and analyzing search engine poisoning of linguistic collisions," *Proc. - IEEE Symp. Secur. Priv.*, vol. 2019-May, pp. 1311–1325, 2019, doi: 10.1109/SP.2019.00025.

[2] A. Kurniawan and L. M. Silalahi, "Analisis Keamanan Jaringan Menggunakan Intrusion Prevention System (IPS) Dengan Metode Traffic Behavior," *Electr. – J. Rekayasa dan Teknol. Elektro*, vol. 17, no. 1, pp. 71–76, 2023.

[3] F. C. Islami, "Analisis Kerentanan Website XYZ Menggunakan Metode Vulnerability Assessment Penetration Testing Dan OWASP WSTG (Studi Kasus: XYZ)," *J. Apl. dan Teor. Ilmu Komput.*, vol. 7, no. 2, pp. 93–99, 2025, doi: 10.17509/jatikom.v7i2.80934.

[4] E. Leka, L. Lamani, A. Aliti, and E. Hoxha, "Web Application Firewall for Detecting and Mitigation of Based DDoS Attacks Using Machine Learning and Blockchain," *TEM J.*, vol. 13, no. 4, pp. 2802–2811, 2024, doi: 10.18421/TEM134-17.

[5] V. Metsis, I. Androutsopoulos, and G. Paliouras, "Spam filtering with Naive Bayes - Which Naive Bayes?," *3rd Conf. Email Anti-Spam - Proceedings, CEAS 2006*, 2006.

[6] P. R. B. Putra, Indriati, and R. S. Perdana, "Klasifikasi Judul Berita Online menggunakan Metode Support Vector Machine (SVM) dengan Seleksi Fitur Chi-square," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, pp. 2132–2141, 2023.

[7] H. Kasdana *et al.*, "Analisis Kerentanan Aplikasi Web terhadap SQL Injection dan XSS Berdasarkan Survei Pengguna dengan Kategori Risiko," *Aisyah J. Informatics Electr. Eng.*, vol. 7, no. 2, pp. 68–75, 2025, [Online]. Available: https://jti.aisyahuniversity.ac.id/index.php/AJIEE

[8] M. Irfa'issurur and B. P. Josaphat, "Machine Learning for Cybersecurity: Web Attack Detection (Brute Force, XSS, SQL Injection)," *Inpr. Indones. J. Pure Appl. Math.*, vol. 7, no. 1, pp. 1–15, 2025, doi: 10.15408/inprime.v7i1.41025.

[9] G. A. Sandag, J. Leopold, and V. F. Ong, "Klasifikasi Malicious Websites Menggunakan Algoritma K-NN Berdasarkan Application Layers dan Network Characteristics," *CogITo Smart J.*, vol. 4, no.

1, pp. 37–45, 2018, doi: 10.31154/cogito.v4i1.100.37-45.

[10] C. Asnawi, D. Hariyadi, U. S. Aesyi, and P. W. Cahyo, "Analisis dan Penanganan Insiden Siber SQL Injection Menggunakan Kerangka NIST SP 800-61R2 dan Algoritma Klusterisasi K-Means," *J. Komtika (Komputasi dan Inform.*, vol. 7, no. 2, pp. 134–144, 2023, doi: 10.31603/komtika.v7i2.10527.

[11] A. Aglasia *et al.*, "Pengenalan wajah pelaku kriminal berbasis sketsa dengan metode segmentasi dan content based image retrieval," vol. 13, no. 3, pp. 1761–1772.

[12] A. Adimas and S. Y. Irianto, "Image Sketch Based Criminal Face Recognition Using Content Based Image Retrieval," *Sci. J. Informatics*, vol. 8, no. 2, pp. 176–182, 2021, doi: 10.15294/sji.v8i2.27865.

[13] R. Rudiansyah, R. Ariyansyah, R. Nanda, and O. Wiranda, "Search Engine Menggunakan Metode Information Retrival," *J. SANTI - Sist. Inf. dan Tek. Inf.*, vol. 2, no. 1, pp. 49–55, 2022, doi: 10.58794/santi.v2i1.68.

[14] Tiya Muthia, N. Nurrahma, and Yudi Eka Putra, "Perbandingan Akurasi Model Pembelajaran Mesin SVM, KNN, Decision Tree, dan Naive Bayes pada Klasifikasi Gangguan Kesehatan Mental," *Electr. J. Rekayasa dan Teknol. Elektro*, vol. 18, no. 3, pp. 363–368, 2024, doi: 10.23960/elc.v18n3.2758.

[15] M. F. As Shidiq and D. Alita, "Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan Svm," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 24–35, 2025, doi: 10.47080/simika.v8i1.3624.

[16] H. Pratama, "Analisis Perbandingan Algoritma Support Vector Machine Dan Artificial Neural Network Dalam Prediksi Dan Klasifikasi Kualitas Pisang," *J. Komput. dan Inform.*, vol. 20, no. 1, pp. 33–48, 2025.

[17] Y. P. Astuti, A. R. Wibowo, E. Kartikadarma, E. R. Subhiyakto, N. A. Sri Winarsih, and M. S. Rohman, "Penerapan Metode Naïve Bayes Classifier Untuk Klasifikasi Sentimen Pada Judul Berita," *LogicLink*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.28918/logiclink.v1i1.7684.